GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 1999, 15:50:46 ; Search time 308.23 Seconds
(without alignments)
12120.749 Million cell updates/sec

Title: US-09-126-945-1
Perfect score: 1894
Sequence: 1 atctgacttcctcccagcac.. .ataaagatactagagaactg 1894

Scoring table: IDENTITY_NUCDX

Searched: 2541579 seqs, 986265752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	556	29.4	606	36	AA662164	AA662164 ns66c07.s
	2	509.2	26.9	514	36	AA662204	AA662204 ns66c07.r
c	3	465.4	24.6	512	50	AI668916	AI668916 wb96d02.x
c	4	374.8	19.8	392	36	AA639821	AA639821 nq84g11.s
	5	348.8	18.4	386	50	AI685592	AI685592 tu20c07.x
c	6	325.2	17.2	366	36	AA662521	AA662521 nu99e07.s
c	7	323.6	17.1	366	50	AI687669	AI687669 tp82g09.x
	8	317.6	16.8	355	34	AA513274	AA513274 nq94e10.s
c	9	308.4	16.3	384	35	AA548598	AA548598 nk54c12.s
	10	291.2	15.4	369	30	AA228304	AA228304 nc38e07.r
	11	259.8	13.7	504	40	AA930371	AA930371 vo45a05.r
c	12	236.2	12.5	441	50	AI686414	AI686414 tu41g05.x
	13	207	10.9	544	46	AI451780	AI451780 mb22f07.y
	14	180.2	9.5	390	22	R75783	R75783 yl21c12.r1
	15	175.2	9.3	245	44	AI305722	AI305722 qw47g09.x
	16	174.8	9.2	216	43	AI223644	AI223644 qx31f05.x
	17	174.8	9.2	223	44	AI270872	AI270872 qw67c10.x
	18	174.8	9.2	222	44	AI271046	AI271046 qw51g09.x
	19	174.8	9.2	244	44	AI275319	AI275319 qw66a10.x
	20	174.8	9.2	232	44	AI275321	AI275321 qw66a12.x
	21	174.8	9.2	224	44	AI284816	AI284816 qw25b04.x
	22	174.8	9.2	199	44	AI289687	AI289687 qw42e08.x
	23	174.8	9.2	220	44	AI289725	AI289725 qw42h11.x
	24	174.8	9.2	218	44	AI305465	AI305465 qw46e08.x
	25	174.8	9.2	218	44	AI305499	AI305499 qw46h09.x
	26	174.8	9.2	231	44	AI305972	AI305972 qw48g01.x
	27	174.8	9.2	231	44	AI305974	AI305974 qw48g03.x
	28	174.8	9.2	222	44	AI305988	AI305988 qw48h09.x
	29	174.8	9.2	212	44	AI306749	AI306749 qw44a10.x
	30	174.8	9.2	220	44	AI306796	AI306796 qw44f02.x
	31	174.8	9.2	215	44	AI306814	AI306814 qw44g09.x
	32	174.8	9.2	227	44	AI308044	AI308044 tb16h08.x
	33	174.8	9.2	242	44	AI308888	AI308888 tb50b07.x
	34	174.8	9.2	242	44	AI309895	AI309895 tb48h06.x
	35	174.8	9.2	232	44	AI310859	AI310859 ta43c12.x
	36	174.8	9.2	228	44	AI311359	AI311359 ta47a08.x
	37	174.8	9.2	219	44	AI311967	AI311967 ta84d11.x
	38	174.8	9.2	217	45	AI340387	AI340387 tb40b10.x
	39	174.8	9.2	217	45	AI340423	AI340423 tb40f02.x
	40	174.8	9.2	217	47	AI473160	AI473160 th54g06.x
	41	174.2	9.2	234	44	AI306750	AI306750 qw44a11.x
	42	173.8	9.2	222	43	AI223570	AI223570 qx30g06.x
	43	173.8	9.2	218	43	AI223720	AI223720 qx32a03.x
	44	173.8	9.2	237	43	AI224320	AI224320 qw99a10.x
	45	173.8	9.2	220	43	AI224659	AI224659 qw97h03.x

ALIGNMENTS

RESULT 1
AA662164/c
LOCUS AA662164 606 bp mRNA EST 03-DEC-1997
DEFINITION ns66c07.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1188588 3' similar to SW:ETS4_DROME P29775 DNA-BINDING PROTEIN D-ETS-4 ;, mRNA sequence.
ACCESSION AA662164
NID g2616255

```
VERSION     AA662204.1  GI:2616295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1392902.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 734   Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham
            High quality sequence stop: 393.
FEATURES                Location/Qualifiers
     source             1. .606
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1188588"
                        /clone_lib="NCI_CGAP_Pr22"
                        /sex="male"
                        /tissue_type="normal prostate"
                        /lab_host="DH10B"
                        /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                        with a modified polylinker; 1st strand cDNA was prepared
                        from normal prostate bulk tissue, and was then primed with
                        a Not I - oligo(dT) primer. Double-stranded cDNA was
                        ligated to Eco RI adaptors (Pharmacia), digested with Not
                        I and cloned into the Not I  and Eco RI sites of the
                        modified pT7T3 vector.  Library is normalized, and was
                        constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      194 a    172 c    204 g    126 t
ORIGIN
```

```
  Query Match             29.4%;  Score 556;  DB 36;  Length 606;
  Best Local Similarity   97.9%;  Pred. No. 7.6e-109;
  Matches  595;  Conservative    0;  Mismatches   10;  Indels    3;  Gaps    3;

Qy    1288 cagacatccagcaaaaaccatcccgccatgaactacgacaagctgagccgctccatccgcca 1347
           ||| |||||||||| |||||||||||||||||||||||||||||| |||||||| ||||
Db     548 CTGGATCCGCAAGAA-CGTCCCAGCATGAACTACGACAAGCTGAGCGGCTCCATCGGCCA 606

Qy    1348 gtatcgccaaaaaggggcatcatccgcgaaagccagacatctcccagcgcctcgtctaccagtt 1407
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db     547 GTATTACAAGAAGGGCATCATCCGGGAAGCCAGACATTTCCCAGCGCCTCGTCTACCAGTT 488

Qy    1408 catgcagcccatctgaatgcctggcccagggcctgaaaccccgccctcaggggcctctctc 1467
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     487 CGTGCAGCCCATCTGAGTGCCTGGCCCAGGGCCTGAAACCCGGCCTCAGGGGCCTCTCTC 428

Qy    1468 ctatcactctctctcaacaaaccctgagatgggggaaaacgggcagtctgctctgctg 1527
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     427 CTGCCTGGCCTGCCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTCTGCTCTGCTG 368

Qy    1528 ctctgacctccagaagcccaagtcagggagggggcaaccaactgccccaggggggatatgg 1587
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     367 CTCTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGATATGG 308
```

```
Qy    1588 gtcctctggggccttcgggaccatggggcaggggtgcttcctcctcaggcccaqctqctc 1647
           |||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Db     307 GTCCTCTGGGG-CTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTC 249

Qy    1648 ccctggaggacagaggggagacaggggctgctccccaacacctgcctctgaccccagcattt 1707
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 CCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTT 189

Qy    1708 ccagagcagagcctacagaagggcagtgactcgacaaaggcccaggcagtccaggcctc 1767
           |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db     188 CCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAGAGCCCAGGCAGTCCAGGCCTC 129

Qy    1768 tctctgctccatccccctgcctcccattctgcaccacacctggcatggtgcaggggaqaca 1827
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 TCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACA 69

Qy    1828 tctgcacccctgagttgggcagccaggagtgcccccg-ggaatggataataaagatacta 1886
           |||||||||||||||| |||||| |||||||||| |||||||||||||||||||||||||
Db      68 TCTGCACCCCTGAGTTGTGCAGCCATGAGTGCCCCCGTGGAATGGATAATAAAGATACTA 9

Qy    1887 gaqaactg 1894
           ||||||||
Db       8 GAGAACTG 1
```

```
RESULT   2
AA662204
LOCUS       AA662204      514 bp    mRNA            EST       03-DEC-1997
DEFINITION  ns66c07 r1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1188588 5'
            similar to SW:ETS4_DROME P29775 DNA-BINDING PROTEIN D-ETS-4 ;, mRNA
            sequence.
ACCESSION   AA662204
NID         g2616295
VERSION     AA662204.1  GI:2616295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1392997.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 734   Std Error: 0.00
            Seq primer: -28ml3 rev1 ET from Amersham.
FEATURES                Location/Qualifiers
     source             1. .514
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1188588"
                        /clone_lib="NCI_CGAP_Pr22"
                        /sex="male"
                        /tissue_type="normal prostate"
                        /lab_host="DH10B"
                        /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                        with a modified polylinker; 1st strand cDNA was prepared
                        from normal prostate bulk tissue, and was then primed with
                        a Not I - oligo(dT) primer. Double-stranded cDNA was
```

ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 138 a 173 c 151 g 82 t
ORIGIN

Query Match 26.9%; Score 509.2; DB 36; Length 514;
Best Local Similarity 99.4%; Pred. No. 6.1e-99;
Matches 511; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy    1241 gaagatcatcttcaaaattgaggactcagcccaggtggcccggctgtggggcatccgcaa 1299
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGAGGGCATCTTCAAAATTGAGGACTCAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAA 60

Qy    1300 gaaccgtcccgccatgaactacgacaagctgagccgctccatccgccagtattacaagaa 1359
           ||| |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      61 GAACCGTCCCGCCATGAACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAA 120

Qy    1360 atgcatcatccagaaagcagaacatctcccagcgcctcgtctaccagttcgtgcaccccat 1419
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGGCATCATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACCCCAT 180

Qy    1420 ctaatgcctgtacccagagcctgaaacccgcctcagggcctctctcctgcctgcctg 1479
           ||| |||||| | |||| || ||||||||||||||||||||||||||||||||||||||||
Db     181 CTGAGTGCCTGGACCAGGGCATGAAACCCGCCTCAGGGGCCTCTCTCCTGCCTGCCCTG 240

Qy    1480 cctcagccaggccctgaggatgggggaaaacgggcagtctgctctgctgctctgaccttcc 1539
           ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTCTGCTCTGCTGCTCTGACCTTCC 300

Qy    1540 aggccctaggctcagggggggggcaaccaactgcccagggggatatgggtcctctgaggc 1599
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGC 360

Qy    1600 ctgcgggcccatcgcgcagggtgcttcctcctcaggcccagctgctcccctggaggaca 1659
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTTCGGGACCATGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACA 420

Qy    1660 aggaaacacaggctgctcccaaacacctgcctctgacccagcatttccagagcagagc 1719
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     421 AAGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGC 480

Qy    1720 ctacagaaagggcagtgactctgacaaaggccacag 1753
           ||||||||||||||||||||||||||||||||||||
Db     481 CTACAGAAGGGCAGTGACTCGACAAAGGCCACAG 514
```

RESULT 1
AI668916/c
LOCUS AI668916 512 bp mRNA EST 14-MAY-1999
DEFINITION ab28b02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE 2313507 3', mRNA sequence.
ACCESSION AI668916
NID g4833690
VERSION AI668916.1 GI:4833690
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 512)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948749.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 441.
FEATURES Location/Qualifiers
source 1..512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2313507"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker, Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101142-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 96 a 142 c 155 g 119 t
ORIGIN

Query Match 24.6%; Score 465.4; DB 50; Length 512;
Best Local Similarity 96.7%; Pred. No. 1.2e-89;
Matches 475; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy    1404 agttcgtgcaccccatctgagtgcctggcccagggcctgaaaccccgccctcagggccc 1463
           |||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     512 AGTTTGTGCACCCCATTTGAGTGCCTGGCCCAGGGCCTGAAACCCCGCCCTCAGGGCCC 453

Qy    1464 tctcctgcctgccctgcctcagccagggcctgagatgggggaaaacgggcagtctgctct 1523
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||| | |
Db     452 TTTCCTGCCTGCCCTGCCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTTTGCTTT 393

Qy    1524 gctgctctgaccttccagagccccaaggtcagggagggggcaaccaactgcccccagggggat 1583
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 GCTGCTTTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGAT 333

Qy    1584 atgggtcctctggggccttcgggaccatgggcagggggtgcttcctcctcaggcccagct 1643
           ||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||| ||
Db     332 ATGGGTCCTTTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAACT 273

Qy    1644 gctccctggaggacagagggagacagggctgctccccaacacctgcctctgacccagc 1703
           |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     272 GCTCCCCTGGAGGACAAAGGGAGACAGGGCTGCTCCCCAACACCTGCCTTTGACCCCAGC 213

Qy    1704 atttccagagcagagcctacagaagggcagtgactcgacaaaggccacaggcagtccagg 1763
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212 ATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGG 153

Qy    1764 cctctctctgctccatcccccctgcctcccattctgcaccacacctggcatggtgcaggga 1823
           ||| | | ||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     152 CCTTTTTTTGCTCCATCCCCCTGCCTCCCATTTTGCACCACACCTGGCATGGTGCAGGGA 93

Qy    1824 gacatctgcacccctgagttgggcagccaggagtgccccccgggaatggataataaagata 1883
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92 GACATTTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATA 33

Qy    1884 ctagagaactg 1894
           |||||||||||
Db      32 CTAGAGAACTG 22
```

```
RESULT   4
AA639821/c
LOCUS       AA639821      392 bp    mRNA            EST       23-OCT-1997
DEFINITION  nq84g11.s1 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1159076 3',
            mRNA sequence.
ACCESSION   AA639821
NID         g2563600
VERSION     AA639821.1  GI:2563600
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 392)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Nov 2, 1998 this sequence version replaced gi:430084

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..392
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1159076"
                     /clone_lib="NCI_CGAP_Co9"
                     /tissue_type="colon tumor RER+"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     RER+ colon tumor, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector. Library is not normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo
                     (Soares4). "
BASE COUNT       65 a    113 c    131 g     83 t
ORIGIN
```

```
  Query Match             19.8%;  Score 374.8;  DB 36;  Length 392;
  Best Local Similarity   99.5%;  Pred. No. 1.8e-70;
  Matches  376;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    1517 ctgctctactgctctgacctccagaggcccaaggtcaggggaggggcaaccaactgcccca 1576
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     478 CTGCTCTGCTGCTCTGAGCTCCCAGAGCCCAAGGTCAGGGAGGGCCAACCAACTGCCCCA 319

Qy    1577 gaaaaatatgaggtcctctaaaacctccagggaccatggggcaggggtgcttcctcctcagg 1636
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     318 GGGGGATATGGGTCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGG 259

Qy    1637 ccccactgctccccctgaaggacagagggagacagggctgctccccaacacctgcctctga 1696
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258 CCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGA 199

Qy    1697 ccccagcattccagagcagagcctacagaagggcagtgactcgacaaaggccacaggca 1756
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198 CCCCAGCATTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCA 139
```

```
Qy    1757 gtccaggcctctctctgctccatcccctgcctcccattctgcaccacacctggcatggt 1816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 GTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCTGGCATGGT 79

Qy    1817 gcagggagacatctgcacccctgagttgggcagccaggagtgccccccgggaaatggataat 1876
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 GCAGGGAGACATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAAT 19

Qy    1877 aaagatactagagaactg 1894
           ||||||||||||||||||
Db      18 AAAGATACTAGAGAACTG 1

RESULT   5
AI685592
LOCUS       AI685592      386 bp    mRNA            EST       27-MAY-1999
DEFINITION  tu20c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PTR5 t3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
NID         g4896886
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188477.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       72 a    140 c    110 g     64 t
ORIGIN
```

```
  Query Match             18.4%;  Score 348.8;  DB 50;  Length 386;
  Best Local Similarity   99.4%;  Pred. No. 6e-65;
```

```
  Matches  85 ;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      [illegible] 349
Db       85 [illegible] 94

Qy      [illegible] 409
Db       95 [illegible] 154

Qy      [illegible] 469
Db      155 AGCGGCATGGGCAGCGGCAGCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCTGCTG 214

Qy      [illegible] 529
Db      215 [illegible] 274

Qy      [illegible] 589
Db      275 [illegible] 334

Qy      [illegible] 641
Db      335 [illegible] 386


RESULT   6
AA662521/c
LOCUS       AA662521      394 bp    mRNA            EST       12-NOV-1997
DEFINITION  [illegible] NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1218852 3',
            mRNA sequence.
ACCESSION   AA662521
NID         g2616612
VERSION     AA662521.1  GI:2616612
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 19, 1997 this sequence version replaced gi:1520374.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.doe.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
FEATURES             Location/Qualifiers
     source          1..366
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1218852"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate. Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       56 a     94 c    110 g    106 t
ORIGIN


  Query Match           17.2%;  Score 325.2;  DB 36;  Length 366;
  Best Local Similarity   99.1%;  Pred. No. 6e-60;
  Matches  327;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    1562 caaccaactgccccaggggatatgggtcctctggggccttcgggaccatggggcagggg 1621
           ||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
Db     366 CAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGCTTTCGGGACCCTGGGGCAGGGG 307

Qy    1622 tgcttcctcctcaggcccagctgtccccctggaggacagaggtggagacagaatgtcccc 1681
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     306 TGCTTCCTCCTCAGGCCCAGCTGTCCCCCTGGAGGACAGAGGTGGAGACAGGGCTGCCCC 247

Qy    1682 aacacctgcctctgaccccagcatttccagagcagagcctacagaaggggcagtgactcga 1741
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     246 AACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGA 187

Qy    1742 caaaggccacaggcagtccaggcctctctctgctccatccccctgcctcccattctgcac 1801
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186 CAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTCCCTTTCTGCAC 127

Qy    1802 cacacctggcatggtgcagggagacatctgcacccctgagttgggcagccaggagtgccc 1861
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 CACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCC 67

Qy    1862 ccgggaatggataataaagatactagagaa 1891
           |||||||||||||||||||||||||||||
Db      66 CCGGGAATGGATAATAAAGATACTAGAGAA 37


RESULT   7
AI687669/c
LOCUS       AI687669      366 bp    mRNA            EST       27-MAY-1999
DEFINITION  tp82g09.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2205856 3',
            mRNA sequence
ACCESSION   AI687669
NID         g4898963
VERSION     AI687669.1  GI:4898963
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Feb 18, 1999 this sequence version replaced gi:4297630.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1..366
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="11 q11-q13.4"
                     /clone="IMAGE:2295856"
                     /clone_lib="NCI_CGAP_Ut3"
                     /tissue_type="poorly-differentiated endometrial
                     adenocarcinoma, 2 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.45 kb. Life Technologies catalog #:
                     11541-018"
BASE COUNT       58 a    104 c    113 g     90 t      1 others
ORIGIN

  Query Match             17.1%;  Score 323.6;  DB 50;  Length 366;
  Best Local Similarity   98.3%;  Pred. No. 1.3e-59;
  Matches  348;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy     1542 agcccaaggtcagggagggcaaccaactgccc-caggggatatgggtcctctggggcc 1600
            |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db      366 AGCCCAAGGTCAGGGAGGGGCAACCAACTG-CCNCAGGGGGATATGGGTCCTCTGGGGCC 308

Qy     1601 ttcgggaccatgggcaggggtgcttcctcctcaggcccagctgctcccctggaggacag 1660
            |||||||||| ||||||||||||||||||||||||||||||||||||||||| |||||||
Db      307 TTCGGGACCCTGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGAAGGACAG 248

Qy     1661 aggggagacagggctgctccccaacacctgcctctgaccccagcatttccagagcagagcc 1720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      247 AGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCC 188

Qy     1721 tacagaagggcagtgactcgacaaaggccacaggcagtccaggcctctctctgctccatc 1780
            |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      187 TACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTTTGCTCCATC 128

Qy     1781 cccctggcctcccatttctgcaccacacctggcatggtgcagggagacatctgcacccctga 1840
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      127 CCCCTGGCTCCCATTTTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGA 68

Qy     1841 gttggggcagccaggagtgcccccgggaatggataataaagatactagagaactg 1894
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       67 GTTGGGGCAGCCAGGAGTGCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 14

RESULT 8
AA513274
LOCUS       AA513274     355 bp    mRNA            EST       19-AUG-1997
DEFINITION  nd94e10.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:942474, mRNA
            sequence.
ACCESSION   AA513274
NID         g2251686
VERSION     AA513274.1  GI:2251686
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            
            Insert Length: 588   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1..355
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:942474"
                     /clone_lib="NCI_CGAP_Pr6"
                     /sex="male"
                     /tissue_type="prostate"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from prostatic
                     intraepithelial neoplasia (low-grade), cDNA made by
                     oligo-dT priming.  Non-directionally cloned.
                     Size-selected on agarose gel, average insert size 600 bp.
                     Reference: Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT       67 a    121 c    109 g     58 t
ORIGIN

  Query Match             16.8%;  Score 317.6;  DB 34;  Length 355;
  Best Local Similarity   98.3%;  Pred. No. 2.4e-58;
  Matches  342;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy     1430 ggcccagggcctgaaacccgccctcaggggcctctctcctgcctgccctgcctcagccag 1489
            || |||||| ||||||||||||||| || |||||||||||||||||||||||||||||
Db        9 GGCACAGGGCATGAAACCCGCCCTCAGTGG-CTCTCTCCTGCCTGCCCTGCCTCAGCCAG 67

Qy     1490 gccctgagatgggggaaaacgggcagtctgctctgctgctctgaccttccagagcccaag 1549
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       68 GCCCTGAGATGGGGGAAAACGGGCAGTCTGCTCTGCTGCTCTGACCTTCCAGAGCCCAAG 127

Qy     1550 gtcagggaggggcaaccaactgccccaggggggatatgggtcctctggggccttcgggacc 1609
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      128 GTCAGGGAGGGGCAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGCCTTCGGGACC 187

Qy     1610 atggggcaggggtgcttcctcctcaggcccagctgctcccctggaggacagagggagaca 1669
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      188 CTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGAGACA 247

Qy     1670 gggctgctccccaacacctgcctctgaccccagcatttccagagcagagcctacagaagg 1729
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248 GGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCCTACAGAAGG 307

Qy     1730 gcagtgactcgacaaaggccacaggcag-tccaggcctctctctgctc 1776
            |||||||||||||||||||||||||||| |||||||||||||||||||
Db      308 GCAGTGACTCGACAAAGGCCACAGGCAGTTCCAGGCCTCTCTCTGCTC 355

RESULT 9
AA548598/c
LOCUS       AA548598     384 bp    mRNA            EST       14-AUG-1997
DEFINITION  nk54c12.s1 NCI_CGAP_Pr7 Homo sapiens cDNA clone IMAGE:1017334, mRNA
            sequence.
ACCESSION   AA548598
NID         g2318880
VERSION     AA548598.1  GI:2318880
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400848.
```

```
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd  ET from Amersham
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1017334"
                     /clone_lib="NCI_CGAP_Pr7"
                     /sex="male"
                     /tissue_type="low-grade prostatic neoplasia"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pAMP10; mRNA made from
                     prostate intraepithelial neoplasia (high-grade), cDNA
                     made by oligo-dT priming.  Non-directionally cloned.
                     Size-selected  on agarose gel, average insert size 600 bp.
                      "
BASE COUNT       93 a    111 c    118 g     92 t
ORIGIN


  Query Match             16.3%;  Score 308.4;  DB 35;  Length 384;
  Best Local Similarity   99.4%;  Pred. No. 2.2e-56;
  Matches  326;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy    1573 ccccaggaaaatatggaatcctctgtggcctttcgggaccatggggcaggggtgcttcctcct 1632
           ||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db     384 TTCAGGGGATATGGGTCCTCTGGGG-CTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCT 326

Qy    1633 caggccaactgctccccctggaggacagaggagacagggctgctccccaacacctgcct 1692
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325 CAGGTCCAGCTGCTCCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCT 266

Qy    1693 ctgaccctagcatttccagagcagagcctacagaagggcagtgactcgacaaaggccaca 1752
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     265 CTGACCCTAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACA 206

Qy    1753 gacagtcaggaaatttctctgctccatcccctgcctcccattctgcaccacacctggca 1812
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 AGCAGTCCAGGGCTCTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCTGGCA 146

Qy    1813 tgatgcaaggaagacatctgcaccctgagttgggcagccaggagtgcccccgggaatgga 1872
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 TGGTGCAGGGAGACATCTGTACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGA 86

Qy    1873 taatuaaaatactaaaauctg 1894
           ||||||||||||||||||||
Db      85 TAATAAAGATACTAGAGAACTG 64


RESULT  10
AA228304
LOCUS       AA228304      369 bp    mRNA            EST       21-AUG-1997
DEFINITION  uc38e07.r1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1010436, mRNA
            sequence.
ACCESSION   AA228304
NID         g1849875
VERSION     AA228304.1  GI:1849875
KEYWORDS    EST.
SOURCE      Human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393445.

            Contact  Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 731   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1010436"
                     /clone_lib="NCI_CGAP_Pr2"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: Not1; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from
                     5,000-10,000 microdissected preneoplastic cells
                     histologically-determined to be prostatic intraepithelial
                     neoplasia 2 (PIN2) cells.  Double-stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG-cloning
                     method (Life Technologies).  Average insert size is 600
                     bp. NOTE: Not directionally cloned.  This library was
                     constructed by  David Krizman."
BASE COUNT       79 a    122 c     98 g     79 t
ORIGIN


  Query Match             15.4%;  Score 291.2;  DB 30;  Length 369;
  Best Local Similarity   99.0%;  Pred. No. 9.7e-53;
  Matches  293;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    1458 ggcctctctcctgcctgcctgcctcaggcaggcccctgagatgggggaaaacctccagtc 1517
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74 GGCCTCTCTCCTGCCTGCCTGCCTCAGCCAGGCCCCTGAGATGGGGGAAAACCTCCAGTC 133

Qy    1518 tgctctgctgctctgaccttccagaggcccaaggtcaggagagggcaaccaactgccccag 1577
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     134 TGCTCTGCTGCTCTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAG 193

Qy    1578 ggggatatgggtcctctggggccttcgggaccatggggcaggggtgcttcctcctcaggc 1637
           |||  ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     194 GGGACTATGGGTCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGC 253

Qy    1638 ccagctgctccccctggaggacagaggagacagggctgctccccaacacctgcctctgac 1697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     254 CCAGCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGAC 313

Qy    1698 cccagcatttccagagcagagcctacagaaggcagtgactcgacaaaggccacag 1753
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     314 CCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAG 369
```

```
RESULT  11
AA930471
LOCUS       AA930471      504 bp    mRNA            EST       23-APR-1998
DEFINITION  vo45a05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone IMAGE:1052816 5' similar to SW:ETS4_DROME P29775 DNA-BINDING
            PROTEIN D-ETS-4 ;, mRNA sequence.
ACCESSION   AA930471
NID         g3079964
VERSION     AA930471.1  GI:3079964
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M , Dietrich,N , Dubuque,T ,
            Geisel,S., Kucaba,T , Lacy,M , Le,M , Martin,J , Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,L., Lennon,G., Soares,B., Wilson,R  and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2287074.

            Contact: Marra M/Mouse EST Project
            WashU HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St  Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:584392
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1..504
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone "IMAGE:1052816"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker, Site_1: EcoRI; Site_2: NotI, Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys   1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG], digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead.  "
BASE COUNT      122 a    149 c    135 g     98 t
ORIGIN


  Query Match          13.7%;  Score 259.8;  DB 40;  Length 504;
  Best Local Similarity   82.5%;  Pred. No. 4.9e-46;
  Matches  511;  Conservative    0;  Mismatches   62;  Indels    4;  Gaps    1;

Qy    1109 gaggaaaaccaaacaaaaccaaaaatggactcatcatgctccgggcagcccatccacctg 1168
           ||||||  |||||| ||   | ||||||||||| || ||||||||||||||||| ||||||
Db       1 GAGGAGGGCTGGACGGATGGTGAGGTGGACTCGTCGTGCTCCGGGCAGCCCATTCACCTG 60

Qy    1169 tgaagattcctcaaaaaattgctactcaagccccacagctatggccgcttcattaggtgg 1228
           |||||||||| || ||  |||| ||||||||||||||||||||||||||||  | |||
Db      61 TGGAGTTCTTGAAAGAACTGCTGCTCAAGCCCCACAGCTATGGCCGCTTCATCCGCTGG 120

Qy    1229 ctcaacaaggagaaaggcatcttcaaaattgaggactcagcccaggtggcccggctgtgg 1288
           |||||||||||||| ||||||||||||||||||||||||||| |||||||||| ||||||

Db     121 CTCAACAAGGAGAAAGGCATCTTCAAAATTGAGGACTCAGCACAGGTGGCCCGACTGTGG 180

Qy    1289 ggcatccgcaaagaaccgtcccgccatgaactacgacaagctgaagccgctccatccgccag 1348
           ||  | |||||||||| || |||||||||| || || || ||||||||||||||||||
Db     181 GGTGTGCGCAAGAACCGGCCAGCCATGAACTATGATAAACTAAGCCGCTCCATCCGCCAG 240

Qy    1349 tattacaagaagggcatcatccggaagccagacatctcccagcgcctcgtctaccagttc 1408
           ||||||||||||||||||| || || || ||||||| ||||||| ||||||| ||
Db     241 TATTACAAGAAGGGCATCATTCGTAAACCCGACATCTCTCAGCGCCTTGTCTACCAGTTT 300

Qy    1409 gtgcaccccatctgagtgcc----tggcccagggcctgaaaccgccctcaggagcctct 1464
           |||||  ||  ||||||  |||     | ||   |||||  |||| ||||    |  ||| ||
Db     301 GTGCATCCAGTCTGAGAGCCACAGAGACCAGAGGCTACAACCTGCCCCAGGCAGCCACT 360

Qy    1465 ctcctgcctgccctgcc 1481
           |||  |   |||  | |
Db     361 CTCTGGTTGGCCTGGTC 377


RESULT  12
AI686414/c
LOCUS       AI686414      441 bp    mRNA            EST       27-MAY-1999
DEFINITION  tu41g05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253656 3',
            mRNA sequence.
ACCESSION   AI686414
NID         g4897708
VERSION     AI686414.1  GI:4897708
KEYWORDS    EST
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189306.

            Contact: Robert Strausberg, Ph D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M  Bento Soares, Ph D
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I M A G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
FEATURES             Location/Qualifiers
     source          1   441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2253656"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M  Fatima Bonaldo.  "
BASE COUNT       96 a     94 c    104 g    143 t      4 others
ORIGIN
```

```
  Query Match             12.5%;  Score 236.2;  DB 50;  Length 441;
  Best Local Similarity   82.0%;  Pred. No. 4.8e-41;
  Matches  296;  Conservative    0;  Mismatches   63;  Indels    2;  Gaps    2;

Qy    1544 gtttccagaagcccaaagtcagggaggggcaaccaactgccccaggggatatgggtcctc 1593
           |  | | ||||||||| ||||||||||||||||| ||||| ||||||||  ||||||
Db       4 -  CCTTTCAAAACCCAAGGTCAAGGAGGGGCAACCAACTTCCCCAAGGGGATATTGGTCCTT 349

Qy    1594 taggacctccgggaccatgggcagggtgcttcctcctcaggcccagctgctcccctgg 1653
           |||||| |||||| | ||||| ||| |||| |||||||||| || || || ||||||
Db     304 TTGGGTTTTGGTACCCTGGGCAAGGGGGCTTTCTCCTCAGGGCCAACTTCTTCCCTGG 289

Qy    1654 aggacaggaaggaaacagggctgctccccaacacctgcctctgaccccagcatttccagag 1713
           ||| | | ||| ||| || || |||||||| |||||||  |  ||||||||||| || |
Db     288 AGGAAAAAAGGAAACAAGGGTGGTCCCCAACCCCTGCCTTTTGCCCCAGCATTTTCAAAA 229

Qy    1714 cagaggcctacaaaaggcaagtgactcga-caaaggccacaggcagtccaggcctctctct 1772
           || |||| || || ||||||| ||    ||||||| |||||| | ||||||| | | |
Db     228 CAAAGCCTTCAAAAAGGCAGTGTCT-TGNCAAAGGCCCCAGGCAATTCAGGCCTTTTTTT 170

Qy    1773 gctccatccccctgcctccccattctgcaccacacctggcatggtgcagggagacatctgc 1832
           |  ||| |||||||| |||| ||| |||||||||| |||| ||| ||| | || | |
Db     169 GGTCCATCCCCCTGCCTCCCATTTTGCCCCACACCTGGGATGGGGCAAGGAAAAATTTTC 110

Qy    1833 acccctgagttgggcagccaggagtgccccccgggaatggataataaagatactagagaac 1892
           |||| ||||||||| |||||||  |||||||||||||||||||||| || ||| |||||
Db     109 TCCCCTTAGTTGGGCAACCAGGAGGGCCCCCGGGAATGGATAATAAAAATTCTAAAGAAC 50

Qy    1893 t 1893

Db      49 T 49


RESULT  13
AI451780
LOCUS       AI451780        544 bp    mRNA            EST       09-MAR-1999
DEFINITION  mb22f07.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:330181 5', mRNA sequence.
ACCESSION   AI451780
NID         g4306393
VERSION     AI451780.1  GI:4306393
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1999 this sequence version replaced gi:2948093.

            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest.watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1..544
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:330181"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTT 3'),
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT      104 a    190 c    153 g     97 t
ORIGIN


  Query Match             10.9%;  Score 207;  DB 46;  Length 544;
  Best Local Similarity   66.4%;  Pred. No. 8e-35;
  Matches  367;  Conservative    0;  Mismatches  165;  Indels   21;  Gaps    4;

Qy      80 cagctccctgcaagcccctcaggttgggccttgccacggtgccagcaggcagccctgggc 139
           | || ||    |   ||| ||| |||||||||| || ||||||||||||  |||||| ||
Db       8 CTGCCCCGGTGAGAACCCCCAGCTTGGGCCTTGTCATGGTGCCAGCAGGTGGCCCTGAGC 67

Qy     140 tgggggtagggggactccctacaggcacgcagccctgagacctcagagggccacccccttga 199
           |  |  |||||| || |||| || |  |||| |||||| |||||||    | | |  | |
Db      68 TTCTGACAGGGGCCTGCCTATAGACCTGCAGGCCTGAGGCCTCAGA----CTCACACTCA 123

Qy     200 gggtggccaggccccagtggccaacctgaagtgctcctctgccaccagccctgctggcc 259
           || |     | |||  |||||| |||| || ||  |||||| | ||||||||||||||
Db     124 AGGGGCAAGAGGCCCTGGTGGCCCACCTAAGAGCCACCTCTGTCCCCAGCCCTGCTGGCC 183

Qy     260 cctggttccgctggccccccagatgcctggctgagacacgccagtggcctcagctgccca 319
           ||         ||| |             | |  ||    | |  ||  ||      ||
Db     184 CCACTGATGTCTGAC----------TGAGACCCAGCAGTGACCCTGAGCTGCCTGCCCAC 233

Qy     320 cacctcttcccggcccctgaagttggcactgcag----cagacagctccctaggcaccaag 375
           |||| ||| || |||||| |||||| |||| |    | |||  |||    | || |||
Db     234 TGCCTCCTCCTGGTCCCTGAGGTTGGCTCTGCCGAGGACGGACGACTCTTCTGACAGCAG 293

Qy     376 gcagctaacagacacagccgccagcccaaacagcagcggcatgggcagcgccaaccccggg 435
           || |||||| ||  |||||| | ||||   || |||   |||||||||| ||||||||  ||
Db     294 GCGGCTAACGGAAGCAGCCCCAAGCCTCCACCGCA---GCATGGGCAGTGCCAGCCCAGG 350

Qy     436 tctgagcagcgtatcccccagccacctcctgctgccccccgacacggtgtcgcggacagg 495
           ||||||| ||| |||||| |   ||| || |||  ||| ||   ||   | || ||||
Db     351 CCTGAGCAACGTGTCCCCCGGTTGCCTGCTACTGTTCCCAGATGTGGCACCACGAACAGG 410

Qy     496 cttggagaaggcggcagcgggggcagtgggtctcgagagacgggactggagtcccagtcc 555
           ||||||||| ||| | || ||| |||| |  |||| | ||| |||||||| ||||
Db     411 GACGGAGAAGGCAGCATCAGGAGCAATGGGCCCTGAGAAGCAGGAATGGAGTCCTAGTCC 470

Qy     556 acccgccacgcccgagcagggcctgtccgccttctacctctcctactttgacatgctgta 615
           ||||||||| || |||||||||||||| || |||||||||| |||||| ||||
Db     471 ACCCGCCACCCCTGAGCAGGGCCTGTCTGCTTTCTACCTCTCTTACTTTAACATGTATCC 530

Qy     616 ccctgaggacagc 628
           |   ||   ||||
Db     531 CGACGATAGCAGC 543


RESULT  14
R75783
LOCUS       R75783          390 bp    mRNA            EST       06-JUN-1995
DEFINITION  yl21c12 r1 Soares breast 2NbHBst Homo sapiens cDNA clone
            IMAGE:158902 5', mRNA sequence.
ACCESSION   R75783
NID         g850465
VERSION     R75783.1  GI:850465
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St  Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 568
            High quality sequence stops: 329
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 568   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 329.
FEATURES              Location/Qualifiers
     source           1. .390
                      /organism="Homo sapiens"
                      /db_xref="GDB:572958"
                      /db_xref="taxon:9606"
                      /map="10"
                      /clone="IMAGE:158902"
                      /clone_lib="Soares breast 2NbHBst"
                      /sex="Female"
                      /dev_stage="adult"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                      Library went through one round of normalization to a Cot =
                      230. Library constructed by Bento Soares and M.Fatima
                      Bonaldo."
BASE COUNT       78 a    118 c    112 g     82 t
ORIGIN
```

```
  Query Match              9.5%;  Score 180.2;  DB 22;  Length 390;
  Best Local Similarity   94.7%;  Pred. No. 3.5e-29;
  Matches  230;  Conservative    0;  Mismatches    8;  Indels    5;  Gaps    4;

Qy    1571 gcaagtcagggaggggcaaccaactg 1601
           |||||||||||||||||||||||||||||
Db     188 GAAGTCAGGGAGGGGCAACCAACTG 214

Qy    1602 ccccaaaaaaataaagtcctctgggcctcaggaccatgggcagggtgctccttcc 1656
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 CCCCAGGGGATATGGGTCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGTGCTTCCTCC 274

Qy    1657 tcaaaaaaaaaaaaatgctcccctcaaaacaaaaaaagacaggactgctccccaacacctg 1716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 TCAGGGGCAGGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTG 320

Qy    1717 cctctaaaaaaaaaaaaatttccaaaaacagaaac-tacagaaaggcagtgactcgacaaag 1775
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db     321 CCTCTGAAGCCCAGAATTTCCAGAGCAGAGCCTTACAGAAGGGCAGTGATTCGACAAAG 380
```

```
Qy    1747 gcc 1749
           |||
Db     369 GGC 371
```

```
RESULT  15
AI305722
LOCUS       AI305722      245 bp    mRNA            EST       03-FEB-1999
DEFINITION  qw47g09.x1 NCI_CGAP_Br13 Homo sapiens cDNA clone IMAGE:1994272 3',
            mRNA sequence
ACCESSION   AI305722
NID         g3990613
VERSION     AI305722.1  GI:3990613
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 245)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 330   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .245
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1994272"
                      /clone_lib="NCI_CGAP_Br13"
                      /sex="female"
                      /tissue_type="breast carcinoma in situ"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                      carcinoma, cDNA made by oligo-dT  priming.  Directionally
                      cloned.  Size-selected on agarose  gel, average insert
                      size 500 bp.  Primary library, non-amplified.  "
BASE COUNT       83 a     66 c     56 g     40 t
ORIGIN
```

```
  Query Match              9.3%;  Score 175.2;  DB 44;  Length 245;
  Best Local Similarity   95.7%;  Pred. No. 3.5e-28;
  Matches  180;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy    1707 tccagagcagagcctacagaagggcagtgactcgacaaaggccacaggcagtccaggcct 1766
           |||| |  || | ||||||||||||||||||||||||||||||||||||||||||||||
Db      18 TGCTGCGAGAAGACGACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCT 77

Qy    1767 ctctctgctccatccccctgcctcccattctgcaccacacctggcatggtgcaggaagac 1826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 CTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGAAGAC 137

Qy    1827 atctgcacccctgagttgggcagccaggagtgccccccaggaatggataataaagatacta 1886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 ATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCCGGAATGGATAATAAAGATACTA 197

Qy    1887 gagaactg 1894
           ||||||||
```

Db 149 GAIAACTS 205

Search Completed: November 20, 1999, 15:57:13
Job Time: 387 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 1999, 15:50:46 ; Search time 308.23 Seconds
(without alignments)
12120.749 Million cell updates/sec

Title: US-09-126-945-1
Perfect score: 1894
Sequence: 1 gtctgacttcctcccagcac.... . .ataaagatactagagaaactg 1894

Scoring table: IDENTITY_NUCLEX

Searched: 2545579 seqs, 995256752 residues

Database: EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	556	29.4	606	36	AA662164	AA662164 ns66c07.s
2	509.2	26.9	514	36	AA662204	AA662204 ns66c07.r
c 3	465.4	24.6	512	50	AI668916	AI668916 wb96d02.x
c 4	374.8	19.8	392	36	AA639821	AA639821 nq84g11.s
5	348.8	18.4	386	50	AI685592	AI685592 tu29c07.x
c 6	325.2	17.2	366	36	AA662521	AA662521 nu99e07.s
c 7	323.6	17.1	366	50	AI687669	AI687669 tp82g09.x
8	317.6	16.8	355	34	AA513274	AA513274 ng94e10.s
c 9	308.4	16.3	384	35	AA548598	AA548598 nk54c12.s
10	291.2	15.4	369	30	AA228304	AA228304 nc38e07.r
11	259.8	13.7	504	40	AA930371	AA930371 vo45a05.r
c 12	236.2	12.5	441	50	AI686414	AI686414 tu41g05.x
13	207	10.9	544	46	AI451780	AI451780 mb22f07.y
14	180.2	9.5	390	22	R75783	R75783 yl21c12.r1
15	175.2	9.3	245	44	AI305722	AI305722 qw47g09.x
16	174.8	9.2	216	43	AI223644	AI223644 qx31f05.x
17	174.8	9.2	223	44	AI270872	AI270872 qw67c10.x
18	174.8	9.2	222	44	AI271046	AI271046 qw51g09.x
19	174.8	9.2	244	44	AI275319	AI275319 qw66a10.x
20	174.8	9.2	232	44	AI275321	AI275321 qw66a12.x
21	174.8	9.2	224	44	AI284816	AI284816 qu25b04.x
22	174.8	9.2	199	44	AI289687	AI289687 qw42e08.x
23	174.8	9.2	220	44	AI289725	AI289725 qw42h11.x
24	174.8	9.2	218	44	AI305465	AI305465 qw46e08.x
25	174.8	9.2	218	44	AI305499	AI305499 qw46h04.x
26	174.8	9.2	231	44	AI305972	AI305972 qw48g01.x
27	174.8	9.2	231	44	AI305974	AI305974 qw48g03.x
28	174.8	9.2	222	44	AI305988	AI305988 qw48h04.x
29	174.8	9.2	212	44	AI306749	AI306749 qw44a10.x
30	174.8	9.2	220	44	AI306796	AI306796 qw44f02.x
31	174.8	9.2	215	44	AI306814	AI306814 qw44g09.x
32	174.8	9.2	227	44	AI308044	AI308044 tb16h08.x
33	174.8	9.2	242	44	AI308888	AI308888 tb50b07.x
34	174.8	9.2	242	44	AI309895	AI309895 tb48h06.x
35	174.8	9.2	232	44	AI310859	AI310859 ta43c12.x
36	174.8	9.2	228	44	AI311359	AI311359 ta47g08.x
37	174.8	9.2	219	44	AI311967	AI311967 ta84d11.x
38	174.8	9.2	217	45	AI340387	AI340387 tb40b10.x
39	174.8	9.2	217	45	AI340423	AI340423 tb40f02.x
40	174.8	9.2	217	47	AI473160	AI473160 th54g06.x
41	174.2	9.2	234	44	AI306750	AI306750 qw44a11.x
42	173.8	9.2	222	43	AI223570	AI223570 qx30g06.x
43	173.8	9.2	218	43	AI223720	AI223720 qx32a03.x
44	173.8	9.2	237	43	AI224320	AI224320 qw99a10.x
45	173.8	9.2	220	43	AI224659	AI224659 qw97h03.x

ALIGNMENTS

RESULT 1
AA662164/c

```
LOCUS       AA662204      514 bp    mRNA            EST       03-DEC-1997
DEFINITION  ns66c07.r1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1188588 3'
            similar to SW:ETS4_DROME P29775 DNA-BINDING PROTEIN D-ETS-4 ;, mRNA
            sequence.
ACCESSION   AA662204
NID         g2616295
VERSION     AA662204.1  GI:2616295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1392997.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 734   Std Error: 0.00
            Seq primer: -28m13 revl ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1188588"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
```

/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 139 a 173 c 151 g 82 t
ORIGIN

Query Match 26.9%; Score 509.2; DB 36; Length 514;
Best Local Similarity 99.4%; Pred. No. 6.1e-99;
Matches 511; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy    1413 aaaaaaaaatcttcaaaaattgaggactcagcccaggtggccccggctgtggggcatccgcaa 1299
           | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGAGGGCATCTTCAAAATTGAGGACTCAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAA 60

Qy    1298 gaaccatcccccccatgaactacgacaagctgagccgctccatccgccagtattacaagaa 1359
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAACCGTCCCCGCCATGAACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAA 120

Qy    1358 aaacatcatccggaaagccagacatctcccagcgcctcgtctaccagttcgtgcaccccat 1419
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GAGCATCATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACCCCAT 180

Qy    1418 ctgagtgcctggcccagggcctgaaacccgccctcaggggcctctctcctgcctgccctg 1479
           | ||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||
Db     181 CTGAGTGCCTGGACCAGGGCATGAAACCCGCCCTCAGGGGCCTCTCTCCTGCCTGCCCTG 240

Qy    1478 cctcagccaggccctgagatgggggaaaacgggcagtctgctctgctgctctgaccttcc 1539
           | ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTCTGCTCTGCTGCTCTGACCTTCC 300

Qy    1538 agagcccaagatcaggggagcaaccaactgcccccagggggatatgggtcctctgggggc 1599
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGAGCCCAAGATCAGGGAGGGGCAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGC 360

Qy    1598 cttcaggaccatgggacagggtgcttcctcctcaggcccagctgctcccctggaggaca 1659
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTTCAGGACCATGGGTCAGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACA 420

Qy    1658 aatgagagcagagtgctgcccaacacctgcctctgaccccagcatttccagagcagagc 1719
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAGGAGAGAGAGATGCTGCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGC 480

Qy    1718 ctacagaaagggcagtgactcgacaaaggccacag 1753
           | |||||||||||||||||||||||||||||||||
Db     481 CTACAGAAGGGCAGTGACTCGACAAAGGCCACAG 514
```

RESULT 2
AI668916/c
LOCUS AI668916 512 bp mRNA EST 14-MAY-1999
DEFINITION xb96b02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE 2313507 3', mRNA sequence.
ACCESSION AI668916
NID g4833690
VERSION AI668916.1 GI:4833690
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 512)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948749.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 441.
FEATURES Location/Qualifiers
source 1..512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2313507"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101954, and 1217929-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 96 a 142 c 155 g 119 t
ORIGIN

Query Match 24.6%; Score 465.4; DB 50; Length 512;
Best Local Similarity 96.7%; Pred. No. 1.2e-89;
Matches 475; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
Qy    1404 agttcgtgcaccccatctgagtgcctggccccagggcctgaaaccccgccctcaggggcctc 1463
           |||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     512 AGTTTGTGCACCCCATTTGAGTGCCTGGCCCAGGGCCTGAAACCCGCCCTCAGGGGCCTC 453

Qy    1464 tctcctgcctgccctgcctcagccaggccctgagatgggggaaaacgggcagtctgctgctct 1523
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db     452 TTTCCTGCCTGCCCTGCCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTCTGCTTT 393

Qy    1524 gctgctctgaccttccagagcccaaggtcagggagggggcaaccaactgccccagggggat 1583
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     392 GCTGCTTTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGAT 333

Qy    1584 atgggtcctctggggccttcgggaccatgggcagggggtgcttcctcctcaggcccagct 1643
           |||||||||| |||||||||||||||| |||||||||||||||||||||||||||||| ||
Db     332 ATGGGTCCTCTTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAACT 273

Qy    1644 gctcccctggaggacagagggagacagggctgctccccaaacacctgcctctgaccccagc 1703
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     272 GCTCCCCTGGAGGACAAAGGGAGACAGGGCTGCTCCCCAACACCTGCCTTTGACCCCAGC 213

Qy    1704 atttccagagcagagcctacagaagggcagtgactcgacaaaggccacaggcagtccagg 1763
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212 ATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGG 153

Qy    1764 cctctctctgctccatcccccctgcctcccattctgcaccacacctaggcatggatgcaggga 1823
           ||| | | |||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     152 CCTTTTTTTGCTCCATCCCCCTGCCTCCCATTTTGCACCACACCTGGCATGGTGCAGGGA 93

Qy    1824 gacatctgcacccctgagttgggcagccaggagtgccccccgggaatggataataaaagata 1883
           ||:|| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92 GACATTTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATA 33
```

```
Qy    1884 ctagaaaactg 1894
           ||||||||||
Db      42 CTAGAGAACTG 22


RESULT   4
AA649821/c
LOCUS       AA649821      392 bp    mRNA            EST       23-OCT-1997
DEFINITION  mp84g11.s1 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1159076 3',
            mRNA sequence.
ACCESSION   AA649821
NID         g2563600
VERSION     AA649821.1  GI:2563600
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 392)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Nov 29, 1998 this sequence version replaced gi:430004.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer:  40ml3 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .392
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1159076"
                     /clone_lib="NCI_CGAP_Co9"
                     /tissue_type="colon tumor PER+"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     REK: colon tumor, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is not normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo
                     (Soares4). "
BASE COUNT       65 a    113 c    131 g     83 t
ORIGIN


  Query Match             19.8%;  Score 374.8;  DB 36;  Length 392;
  Best Local Similarity   99.5%;  Pred. No. 1.8e-70;
  Matches  376;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    1517 ctgctctgctgctctgacctccagaagcccaaggtcagggaggggcaaccaactgcccca 1576
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     319 CTGCTCTGCTGCTCTGACCTCCAGAGCCCAAGGTCAGGGAGGGCAACCAACTGCCCCA 378

Qy    1577 cgaggatatgagtgtctctgaggcttcgggaccatggggcaggggtgcttcctcctcagg 1636
           |||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||
Db     379 CGAGGATATGAGTGCCTCTGAGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGG 438

Qy    1637 ccccgctgctccccaacacctgcctctga 1696
           |||||||||||||||||||||||||||||
Db     439 CCCCGCTGCTCCCCAACACCTGCCTCTGA 468
```

```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258 CCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGA 199

Qy    1697 ccccagcatttccagagcagagcctacagaagggcagtgactgacaaagggccacagaca 1756
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198 CCCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTGGACAAAGGCCACAGGCA 139

Qy    1757 gtccaggcctctctctgctccatccccctgcctcccattctgcaccacacctgggcatggt 1816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     138 GTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCTGGGCATGGT 79

Qy    1817 gcagggagacatctgcacccctgaqttgggcaqccaqqaqtgccccccqqqaatqqataat 1876
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 GCAGGGAGACATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCCGGGAATGGATAAT 19

Qy    1877 aaagatactagagaactg 1894
           ||||||||||||||||||
Db      18 AAAGATACTAGAGAACTG 1


RESULT   5
AI685592
LOCUS       AI685592      386 bp    mRNA            EST       27-MAY-1999
DEFINITION  tu20c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PTR5.t3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
NID         g4896886
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188477.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 380
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
```

BASE COUNT 72 a 140 c 110 g 64 t
ORIGIN

Query Match 18.4%; Score 348.8; DB 50; Length 386;
Best Local Similarity 99.4%; Pred. No. 6e-65;
Matches 350; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 249 ctgaaacatggcagtggcctcagctgcccacacctcttcccggcccctgaagttggcact 349
Db 46 TGGGCATAGGCCAGTGGCCTCAGCTGCCCACACCTCTTCCCGGCCCCTGAAGTTGGCACT 94

Qy 350 aaaaagacaagctcctgggcaccaggcagctaacagacacagccgccagcccaaacagc 409
Db 95 GAAGTAGATAGTCCCCTGGGCAGGCTAACAGACACAGCCGCCAGCCCAAACAGC 154

Qy 410 aaggcatgggcaggcagccagcccgggtctgaggcagcgtatccccagccacctcctgctg 469
Db 155 AGGGCATGGGCAGCCTAGCCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCTGCTG 214

Qy 470 ccccccgacaggatgtgcgggacaggctggagaaggcggcagcgggggcagtgggtctc 529
Db 215 CCCCCCGACAGGATGTCCGGGACAGGCTTGGAGAAGGCGGCAGCGGGGGCAGTGGGTCTC 274

Qy 530 gagagacgggactggagtccagtccacccgccacgcccgagcagggcctgtccgccttc 589
Db 275 GAGAGACGGGACTGGAGTCCCAGTCCACCCGCCACGCCCGAGCAGGGCCTGTCCGCCTTC 334

Qy 590 tgcctctctgctttgagatgctgtacctgaggacagcagctgggcagcca 641
Db 335 TGCCTCTCTGCTTTGAGATGCTGTACCTGAGGACAGCAGCTGGGCAGCCA 386

RESULT 6
AA662521/c
LOCUS AA662521 366 bp mRNA EST 12-NOV-1997
DEFINITION [illegible] mRNA sequence.
ACCESSION AA662521
NID g2616612
VERSION AA662521.1 GI:2616612
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota, Metazoa, Chordata, Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 19, 1997 this sequence version replaced gi:1520374.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: 40m13 fwd. ET from Amersham.
FEATURES Location/Qualifiers
source 1..366
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1218852"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 56 a 94 c 110 g 106 t
ORIGIN

Query Match 17.2%; Score 325.2; DB 36; Length 366;
Best Local Similarity 99.1%; Pred. No. 6e-60;
Matches 327; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1562 caacaactgcccccaggggcatatgggtcctctgggggcttcgggacctgggggcaggggg 1621
Db 366 CAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGCTTTCGGGACCCTGGGGCAGGGG 307

Qy 1622 tgcttcctcctcaggcccagctgctcccctggaggacagaggagagcagggctgctcccc 1681
Db 306 TGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCC 247

Qy 1682 aacacctgcctctgacccccagcatttccagaggcagaggcctacagaaggggcagtgactcga 1741
Db 246 AACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGGCCTACAGAAGGGCAGTGACTCGA 187

Qy 1742 caaaggcccacaggcagtccaggcctctctctgctccatccccctgcctcccattctgcac 1801
Db 186 CAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTCCCTTTCTGCAC 127

Qy 1802 cacacctggcatggtgcagggagacatctgcacccctgagttgggcagccaggagtcccc 1861
Db 126 CACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAGTTGGGCAGCCAGGAGTCCCC 67

Qy 1862 ccgggaatggataataaagatactagagaa 1891
Db 66 CCGGGAATGGATAATAAAGATACTAGAGAA 37

RESULT 7
AI687669/c
LOCUS AI687669 366 bp mRNA EST 27-MAY-1999
DEFINITION tp82g09.x1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:2205856 3',
mRNA sequence.
ACCESSION AI687669
NID g4898963
VERSION AI687669.1 GI:4898963
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Mammalia,
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Feb 18, 1999 this sequence version replaced gi:4297630.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

```
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer:  40UP from Gibco
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="11 q11-q13.4"
                     /clone="IMAGE:2205856"
                     /clone_lib="NCI_CGAP_Ut3"
                     /tissue_type="poorly-differentiated endometrial
                     adenocarcinoma, 2 pooled tumors"
                     /lab_host="DH10B"
                     /note "Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.45 kb. Life Technologies catalog #:
                     11941-018"
BASE COUNT       58 a    104 c    113 g     90 t      1 others
ORIGIN


  Query Match             17.1%;  Score 323.6;  DB 50;  Length 366;
  Best Local Similarity   98.8%;  Pred. No. 1.3e-59;
  Matches  348;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy    1542 gccccaaaatcaaaaaaaaacaaccaactgccc-caggggatatgggtcctctggggcc 1600
           |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db     366 AGCCCAAGGTCAGGGAGGGGCAACCAACTG-CCNCAGGGGGATATGGGTCCTCTGGGGCC 308

Qy    1601 ttcgaaaccatgaaaacaaaagtgcttcctcctcaggcccagctgctcccctggaggacag 1660
           |||||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     307 TTCGGGACGGCTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGAAGGACAG 248

Qy    1661 aaaaaatacaaaactgctccccaacacctgcctctgaccccagcatttccagagcagagcc 1720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247 AGGGAGACAGGGGCTGTGGCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCC 188

Qy    1721 tacaaaaaaaaaaaatgactcctacaaaaggccacaggcagtccaggcctctctctgctccatc 1780
           ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     187 TACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTTTGCTCCATC 128

Qy    1781 cccctgcctccctttcgcaccacacctggcatggtgcagggagacatctgcaccccctga 1840
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     127 CCCCTGCCTCCCATTTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGA 68

Qy    1841 gtttgggaaaaccaaaagtgccccccgggaatggataataaagatactagagaactg 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67 GTTTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 14


RESULT   8
AA513274
LOCUS       AA513274       355 bp    mRNA            EST       19-AUG-1997
DEFINITION  nd94c10.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:942474, mRNA
            sequence.
ACCESSION   AA513274
NID         g2251686
VERSION     AA513274.1  GI:2251686
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 588   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1. .355
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:942474"
                     /clone_lib="NCI_CGAP_Pr6"
                     /sex="male"
                     /tissue_type="prostate"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from prostatic
                     intraepithelial neoplasia (low-grade). cDNA made by
                     oligo-dT priming.  Non-directionally cloned.
                     Size-selected on agarose gel, average insert size 600 bp.
                     Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT       67 a    121 c    109 g     58 t
ORIGIN


  Query Match             16.8%;  Score 317.6;  DB 34;  Length 355;
  Best Local Similarity   98.3%;  Pred. No. 2.4e-58;
  Matches  342;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy    1430 ggcccaggggcctgaaaacccgccctcaggggcctctctcctgcctgccctgcctcagccag 1489
           ||| |||||| |||||||||||||||| || ||||||||||||||||||||||||||||||
Db       9 GGCACAGGGCATGAAACCCGCCCTCAGTGG-CTCTCTCCTGCCTGCCCTGCCTCAGCCAG 67

Qy    1490 gccctgagatgggggaaaacgggcagtctgctctgctgctctgaccttccagagcccaag 1549
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 GCCCTGAGATGGGGGAAAACGGGCAGTCTGCTCTGCTGCTCTGACCTTCCAGAGCCCAAG 127

Qy    1550 gtcagggagggggcaaccaactgccccaggggatatgggtcctctgggcctcagggacc 1609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 GTCAGGGAGGGGCAACCAACTGCCCCAGGGGATATGGGTCCTCTGGGCCTCAGGGACC 187

Qy    1610 atggggcaggggtgcttcctcctcaggcccagctgctcccctggaggacagagggagaca 1669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188 CTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGAGACA 247

Qy    1670 gggctgctccccaacacctgcctctgaccccagcatttccagagcagagcctacagaaag 1729
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 GGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCCTACAGAAGG 307

Qy    1730 gcagtgactcgacaaaggccacaggcag-tccaggcctctctctgctc 1776
           |||||||||||||||||||||||||||| ||||||||||||||||||
Db     308 GCAGTGACTCGACAAAGGCCACAGGCAGTTCCAGGCCTCTCTCTGCTC 355


RESULT   9
AA548598/c
LOCUS       AA548598       384 bp    mRNA            EST       14-AUG-1997
DEFINITION  nk54c12.s1 NCI_CGAP_Pr7 Homo sapiens cDNA clone IMAGE:1017334, mRNA
            sequence.
ACCESSION   AA548598
NID         g2318880
VERSION     AA548598.1  GI:2318880
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 384)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400848.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..384
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1017334"
                     /clone_lib="NCI_CGAP_Pr7"
                     /sex="male"
                     /tissue_type="low-grade prostatic neoplasia"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pAMP10; mRNA made from
                     prostate intraepithelial neoplasia (high-grade), cDNA
                     made by oligo-dT priming.  Non-directionally cloned.
                     Size-selected on agarose gel, average insert size 600 bp.
                     "
BASE COUNT       63 a    111 c    118 g     92 t
ORIGIN


  Query Match             16.3%;  Score 308.4;  DB 35;  Length 384;
  Best Local Similarity   99.4%;  Pred. No. 2.2e-56;
  Matches  321;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy    1573 ...cagggguatatgggtcctctgggggccttcgggaccatggggcaggggtgcttcctcct 1632
           ||| | ||||||||| |||||||| |||||||||| ||||||||||||||||||||||||
Db     384 TCCAGGGGGATATGGGTCCTCTGGGG-CTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCT 326

Qy    1633 ...|.|.|||||||.|.|| |||||||cagaggggagacagggctgctccccaacaacctgct 1692
               | |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325 TAGGTTAGTTGTTTTTTTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCT 266

Qy    1693 ctgaccccaggcatttccagagcagaggcctacagaagggcagtgactcgacaaaggccaca 1752
           |||| ||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     265 CTGAGCCCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACA 206

Qy    1753 ...|||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 1812
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 AGCAGCCAGGCCTCTGCTGCTCCACACCCCCTGCCTCCCATTCTGCACCACACCTGGCA 146

Qy    1813 tggtgcaggagaagacatctgcacccctgagttgggcagccaggagtgcccccgggaatgga 1872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 TGGTGCAGGGAGACATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGA 86

Qy    1873 taatbaaiatactatagaaactg 1894
           ||||||||||||||||||||||
Db      85 TAATAAAGATACTATAGAAATG 64


RESULT  1
AA228304
LOCUS       AA228304      369 bp    mRNA            EST       21-AUG-1997
DEFINITION  ng38e07.r1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1010436, mRNA
            sequence.
ACCESSION   AA228304

NID         g1849875
VERSION     AA228304.1  GI:1849875
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393445.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 731   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1..369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1010436"
                     /clone_lib="NCI_CGAP_Pr2"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from
                     5,000-10,000 microdissected preneoplastic cells
                     histologically-determined to be prostatic intraepithelial
                     neoplasia 2 (PIN2) cells.  Double-stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the TOPO cloning
                     method (Life Technologies).  Average insert size is 600
                     bp. NOTE: Not directionally cloned.  This library was
                     constructed by David Krizman."
BASE COUNT       70 a    122 c     98 g     79 t
ORIGIN


  Query Match             15.4%;  Score 291.2;  DB 30;  Length 369;
  Best Local Similarity   99.0%;  Pred. No. 9.7e-53;
  Matches  293;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    1458 ggcctctctcctgcctgccctgcctcagccaggcccctgagatgggggaaaacgggcagtc 1517
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74 GGCCTCTCTCCTGCCTGCCCTGCCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTC 133

Qy    1518 tgctctgctgctctgaccttccagagcccaaggtcaggaaggagcaaccaactgcccccag 1577
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     134 TGCTCTGCTGCTCTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAG 193

Qy    1578 ggggatatgggtcctctggggccttcgggaccatggggcaggggtgcttcctcctcaggc 1637
           ||| ||||||||||||||||||||||||||| ||||||||||||||||||| ||| |||
Db     194 GGGACTATGGGTCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGC 253

Qy    1638 ccagctgctccccctggaggacagaaggagacagggctgctccccaacacctgcctctgac 1697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     254 CCAGCTGCTCCCCTGGAGGACAGAGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGAC 313
```

```
Qy    1048 ccccagcatttccagagcagagcctacagaagggcagtgactcgacaaaggccacag 1753
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      424 CCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAG 369


RESULT 11
AA930471/c
LOCUS       AA930471      504 bp    mRNA            EST       23-APR-1998
DEFINITION  vo45a05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone IMAGE:1052816 5' similar to SW:ETS4_DROME P29775 DNA-BINDING
            PROTEIN D-ETS-4 ;, mRNA sequence.
ACCESSION   AA930471
NID         g3079964
VERSION     AA930471.1  GI:3079964
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Stoptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:2287074.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MCI:584392
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1052816"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG], digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector. Library constructed by Bob Barstead. "
BASE COUNT      122 a    149 c    135 g     98 t
ORIGIN


  Query Match          13.7%;  Score 259.8;  DB 40;  Length 504;
  Best Local Similarity   82.5%;  Pred. No. 4.9e-46;
  Matches  411;  Conservative    0;  Mismatches   62;  Indels    4;  Gaps    1;

Qy    1169 tggcagttcctcaaggagttgctactcaagccccacagctatgggcgcttcattaggtgg 1228
           ||||||||||| || ||  |||| ||||||||||||||||||||||||||||  | |||
Db      61 TGGCAGTTCCTGAAAGAACTGCTGCTCAAGCCCCACAGCTATGGCGGCTTCATCGGCTGG 120

Qy    1229 ctcaacaaggagaagggcatcttcaaaattgaggactcagcccagatgtgcccgactgtgg 1288
           |||||||||||||| ||||||||||||||||||||||||| |||||||||| ||||||
Db     121 CTCAACAAGGAGAAAGGCATCTTCAAAATTGAGGACTCAGCACAGGTGGGCCGACTGTGG 180

Qy    1289 ggcatccgcaagaaccgtccgccatgaactacgacaagctgaggcggctccatccgccag 1348
           ||  | |||||||||| || |||||||||| || || || ||||||||||||||||||
Db     181 GGTGTGGGCAAGAACCGGCCAGCCATGAACTATGATAAACTAAGCCGCTCCATCCGCCAG 240

Qy    1349 tattacaagaagggcatcatccggaagccagacatctcccagcggcctgtctaccagttc 1408
           |||||||||||||||||||| || || || |||||||| |||||||| |||||||| ||
Db     241 TATTACAAGAAGGGCATCATTCGTAAACCCGACATCTCTCAGCGGCTTGTCTACCAATTT 300

Qy    1409 gtgcaccccatctgagtgcc----tggcccagggcctgaaaccccgcctcaggggcctct 1464
           ||||| ||  |||||| |||     | ||   |||||  |||| |||   |  ||| ||
Db     301 GTGCATCCAGTCTGAGAGCCACAGAGACCAGAGGCCTACAACCTGCCCCAGGCAGCCACT 360

Qy    1465 ctcctgcctgccctgcc 1481
           |||  |   |||  | |
Db     361 CTCTGGTTGGCCTGGTC 377


RESULT 12
AI686414/c
LOCUS       AI686414      441 bp    mRNA            EST       27-MAY-1999
DEFINITION  tu41g05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253656 3',
            mRNA sequence.
ACCESSION   AI686414
NID         g4897708
VERSION     AI686414.1  GI:4897708
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189306.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2253656"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
```

of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 96 a 94 c 104 g 143 t 4 others
ORIGIN

Query Match 12.5%; Score 236.2; DB 50; Length 441;
Best Local Similarity 82.0%; Pred. No. 4.8e-41;
Matches 296; Conservative 0; Mismatches 63; Indels 2; Gaps 2;

Qy 1544 gcttccagagccaaggtcagggagggcaacaaactgcccagggggatatggtcctc 1593
Db 48 CTTTCAAAACCCAAGTCAAGGAGGGGCAACCAACTTCCCCAAGGGGATATTGGTCCTT 349

Qy 1594 tgaggcttcgggaccatgggcagggtgcttcctcctcaggcccagctgctcccctgg 1653
Db 348 TGGGCTTTTGGGACCCTTGGGCAAGGGGCTTTCTCCTCAGGGCCAACTTCTTCCCTGG 289

Qy 1654 agtacagaggggagacaggggtgctcccaacacctgcctctgaccccagcatttccagag 1713
Db 288 AGGAAAAAAGGAAACAAGGGTGGTCCCCAACCCCTGCCTTTTGCCCCAGCATTTTCAAAA 229

Qy 1714 catagcctacagaaaggcagtgactcga-caaaggccacaggcagtccaggcctctctct 1772
Db 228 CAAACCTTCAAAAA3GCAGTGTCT-TGNCAAAGGCCCCAGGCAATTCAGGCCTTTTTTT 170

Qy 1773 gctccatccccctgcctcccattctgcaccacacctggcatggtgcagggagacatctgc 1832
Db 169 GGTCCATCCCCCTGCCTCCCATTTTGCCCCACACCTGGGATGGGGCAAGGAAAAATTTTC 110

Qy 1833 acccctgagttuuucagccaggagtgccccccgggaatggataataaagatactagagaac 1892
Db 109 TCCCCTTAGTTGGGTAACCAGGAGGGCCCCCGGGAATGGATAATAAAAATTCTAAAGAAC 50

Qy 1893 t 1893
Db 49 T 49

RESULT 13
AI451780
LOCUS AI451780 544 bp mRNA EST 09-MAR-1999
DEFINITION mb02f07.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:330181 5', mRNA sequence
ACCESSION AI451780
NID g4306393
VERSION AI451780.1 GI:4306393
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 544)
AUTHORS Marra M., Hillier L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons M., Pape,D., Harvey N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT On Mar 10, 1999 this sequence version replaced gi:2948093.

Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 462.
FEATURES Location/Qualifiers
source 1..544
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:330181"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT7T3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
BASE COUNT 104 a 190 c 153 g 97 t
ORIGIN

Query Match 10.9%; Score 207; DB 46; Length 544;
Best Local Similarity 66.4%; Pred. No. 8e-35;
Matches 367; Conservative 0; Mismatches 165; Indels 21; Gaps 4;

Qy 80 cagctccctgcaagccccctcaggttgggccttgccacggtgccagcagggcaccctgagc 139
Db 8 CTGCCCCGGTGAGAACCCCCAGCTTGGGCCTTGTCATGGTGCCAGCAGGTGGCCCTGAGC 67

Qy 140 tgggggtaggggactccctacaggcacgcagccctgagacctcagagggccaccccttga 199
Db 68 TTCTGACAGGGGCCTGCCTATAGACCTGCAGGCCTGAGGCCTCAGA----CTCACACTCA 123

Qy 200 gggtggccaggccccagtggccaacctgagtgctgcctctgccaccagccctactgagcc 259
Db 124 AGGGGCAAGAGGCCCTGGTGGCCCACCTAAGAGCCACCTCTGTCCCCAGCCCTGCTGGCC 183

Qy 260 cctggttccgctggccccccagatgcctggctgagacacgccagtggcctcagctgccca 319
Db 184 CCACTGATGTCTGAC----------TGAGACCCAGCAGTGACCCTGAGCTGCCTGCCCAC 233

Qy 320 cacctcttcccggccccctgaagttggcactgcag----cagacagctccctgggcaccag 375
Db 234 TGCCTCCTCCTGGTCCCTGAGGTTGGCTCTGCCGAGGACGGACGACTCTTCTGACACCAG 293

Qy 376 gcagctaacagacacagccgccagcccaaacagcagcggcatgggcagcccagcccgggg 435
Db 294 GCGGCTAACGGAAGCAGCCCCAAGCCTCCACCGCA---GCATGGGCAGTGCCAGCCCAGG 350

Qy 436 tctgagcagcgtatccccccagccacctcctgctgcccccccgacacggtgtcccgggacagg 495
Db 351 CCTGAGCAACGTGTCCCCCGGTTGCCTGCTACTGTTCCCAGATGTGGCACCACGAACAGG 410

Qy 496 cttggagaaggcggcagcgggggcagtgggtctcgaggagacgggactgggatcccagtcc 555
Db 411 GACGGAGAAGGCAGCATCAGGAGCAATGGGCCCTGAGAAGCAGGAATGGAGTCCTAGTCC 470

Qy 556 acccgccacgcccgagcagggcctgtccgccttctaccctcctcctacttttgacatgctgta 615
Db 471 ACCCGCCACCCCTGAGCAGGGCCTGTCTGCTTTCTACCTCTCTTACTTTAACATGTATCC 530

Qy 616 ccctgaggacagc 628
Db 531 CGACGATAGCAGC 543

RESULT 14
R75783

```
LOCUS       R75784          390 bp    mRNA            EST       06-JUN-1995
DEFINITION  yi21c12.r1 Soares breast 2NbHBst Homo sapiens cDNA clone
            IMAGE:158902 5', mRNA sequence.
ACCESSION   R75784
NID         g850465
VERSION     R75784.1  GI:850465
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 568
            High quality sequence stops: 329
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 568   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 329.
FEATURES             Location/Qualifiers
     source          1..390
                     /organism="Homo sapiens"
                     /db_xref="GDB:572958"
                     /db_xref="taxon:9606"
                     /map="10"
                     /clone="IMAGE:158902"
                     /clone_lib="Soares breast 2NbHBst"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                     Library went through one round of normalization to a Cot =
                     240. Library constructed by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       78 a    118 c    112 g     82 t
ORIGIN
```

```
  Query Match             9.5%;  Score 180.2;  DB 22;  Length 390;
  Best Local Similarity   94.7%;  Pred. No. 3.5e-29;
  Matches  238;  Conservative    8;  Mismatches    8;  Indels    5;  Gaps    4;

Qy    1512 gcaatctgctctgctgctctgaccttccagagcccaaggtcagggagggggcaaccaactg 1571
           |||   ||| |||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     120 GAAGTGCACTCTGCTGCTCTGACGTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTG 188

Qy    1572 ccccaaaaaatatgtgatctctgaggccttcagggaccatgggcagggggtgcttcctcc 1631
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     189 CCCAGGGGGATATGGGTCCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCC 248

Qy    1632 tcaaaccccaggtgctcccgtag--aggacagagggagacagggctgctccccaacacctg 1689
           |||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
```

```
Db     249 TCAGGCCCAGCTGCTCCCCTGGGAGGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTG 308

Qy    1690 gctc-tgaccccagca-tttccagagcagagcc-tacagaagggcagtgactcgacaaaq 1746
           |||| |||||||||| ||||||||||||||||| ||||||||||||||| |||||||||
Db     309 CCTCTTGACCCCAGCATTTTCCAGAGCAGAGCCTTACAGAAGGGCAGTGATTCGACAAAG 368

Qy    1747 gcc 1749
           | |
Db     369 GGC 371


RESULT  15
AI305722
LOCUS       AI305722       245 bp    mRNA            EST       03-FEB-1999
DEFINITION  qw47g09 x1 NCI_CGAP_Br13 Homo sapiens cDNA clone IMAGE:1994272 3',
            mRNA sequence.
ACCESSION   AI305722
NID         g3990613
VERSION     AI305722.1  GI:3990613
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 245)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 330   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1   245
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1994272"
                     /clone_lib="NCI_CGAP_Br13"
                     /sex="female"
                     /tissue_type="breast carcinoma in situ"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     carcinoma, cDNA made by oligo-dT  priming.  Directionally
                     cloned.  Size-selected on agarose  gel, average insert
                     size 500 bp.  Primary library, non-amplified.  "
BASE COUNT       83 a     66 c     56 g     40 t
ORIGIN
```

```
  Query Match             9.3%;  Score 175.2;  DB 44;  Length 245;
  Best Local Similarity   95.7%;  Pred. No. 3.5e-28;
  Matches  180;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy    1707 tccagagcagagcctacagaagggcagtgactcgacaaagggccacaggcagtccaggcct 1766
           | | | |   || | ||||||||||||||||||||||||||||||||||||||||||||||
Db      18 TGCTGCGAGAAGACGACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCT 77

Qy    1767 ctctctgctccatccccctgcctcccattctgcaccacacctggcatggtgcaggagagac 1826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 CTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGAGAGAC 137
```

```
Qy    1827 atctgcagccctgaagttggagcagccaggagtgccccccgggaatggataataaagatacta 1886
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     148 ATCTGCAGCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTA 197

Qy    1887 gagaactg 1894
           ||||||||
Db     198 GAGAACTG 205


Search completed: November 20, 1999, 15:57:13
Job time: 387 sec

                          GenCore version 4.5
                 Copyright (c) 1993 - 1998  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         November 20, 1999, 15:50:46 ; Search time 308.23 Seconds
                                        (without alignments)
                                        12120.749 Million cell updates/sec

Title:          US-09-126-945-1
Perfect score:  1894
Sequence:       1 atctgaattcctcccagcac.........ataaagatactagagaactg 1894

Scoring table:  IDENTITY_NUCDX

Searched:       2546578 seqs, 986266752 residues

Database :      EST:*
                1:  em_est1:*
                2:  em_est2:*
                3:  em_est3:*
                4:  em_est4:*
                5:  em_est5:*
                6:  em_est6:*
                7:  em_est7:*
                8:  em_est8:*
                9:  em_est9:*
                10:  em_est10:*
                11:  em_est11:*
                12:  em_est12:*
                13:  em_est13:*
                14:  em_est14:*
                15:  em_est15:*
                16:  em_est16:*
                17:  em_est17:*
                18:  em_est18:*
                19:  em_est19:*
                20:  gb_est1:*
                21:  gb_est2:*
                22:  gb_est3:*
                23:  gb_est4:*
                24:  gb_est5:*
                25:  gb_est6:*
                26:  gb_est7:*
                27:  gb_est8:*
                28:  gb_est9:*
                29:  gb_est10:*
                30:  gb_est11:*
                31:  gb_est12:*
                32:  gb_est13:*
                33:  gb_est14:*
                34:  gb_est15:*
                35:  gb_est16:*
                36:  gb_est17:*
                37:  gb_est18:*
                38:  gb_est19:*
                39:  gb_est20:*
                40:  gb_est21:*
                41:  gb_est22:*
                42:  gb_est23:*
                43:  gb_est24:*
                44:  gb_est25:*
                45:  gb_est26:*
                46:  gb_est27:*
                47:  gb_est28:*
                48:  gb_est29:*
                49:  gb_est30:*
                50:  gb_est31:*
                51:  gb_est32:*
                52:  em_est20:*
                53:  em_est21:*
                54:  em_est22:*
                55:  em_est23:*
                56:  em_est24:*
                57:  em_est25:*
                58:  em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	556	29.4	606	36	AA662164	ns66c07.s
	2	509.2	26.9	514	36	AA662204	AA662204 ns66c07.r
c	3	465.4	24.6	512	50	AI668916	AI668916 wb96d02.x
c	4	374.8	19.8	392	36	AA639821	AA639821 nq84g11.s
	5	348.8	18.4	386	50	AI685592	AI685592 tu20c07.x
c	6	325.2	17.2	366	36	AA662521	AA662521 nu99e07.s
c	7	323.6	17.1	366	50	AI687669	AI687669 tp82g09.x
	8	317.6	16.8	355	34	AA513274	AA513274 nq94e10.s
c	9	308.4	16.3	384	35	AA548598	AA548598 nk54c12.s
	10	291.2	15.4	369	30	AA228304	AA228304 nc38e07.r
	11	259.8	13.7	504	40	AA930371	AA930371 vo45a05.r
c	12	236.2	12.5	441	50	AI686414	AI686414 tu41g05.x
	13	207	10.9	544	46	AI451780	AI451780 mb22f07.y
	14	180.2	9.5	390	22	R75783	R75783 yl21c12.r1
	15	175.2	9.3	245	44	AI305722	AI305722 qw47g09.x
	16	174.8	9.2	216	43	AI223644	AI223644 qx31f05.x
	17	174.8	9.2	223	44	AI270872	AI270872 qw67c10.x
	18	174.8	9.2	222	44	AI271046	AI271046 qw51g09.x
	19	174.8	9.2	244	44	AI275319	AI275319 qw66a10.x
	20	174.8	9.2	232	44	AI275321	AI275321 qw66a12.x
	21	174.8	9.2	224	44	AI284816	AI284816 qw25b04.x
	22	174.8	9.2	199	44	AI289687	AI289687 qw42e08.x
	23	174.8	9.2	220	44	AI289725	AI289725 qw42h11.x
	24	174.8	9.2	218	44	AI305465	AI305465 qw46e08.x
	25	174.8	9.2	218	44	AI305499	AI305499 qw46h09.x
	26	174.8	9.2	231	44	AI305972	AI305972 qw48g01.x
	27	174.8	9.2	231	44	AI305974	AI305974 qw48g03.x
	28	174.8	9.2	222	44	AI305988	AI305988 qw48h09.x
	29	174.8	9.2	212	44	AI306749	AI306749 qw44a10.x
	30	174.8	9.2	220	44	AI306796	AI306796 qw44f02.x
	31	174.8	9.2	215	44	AI306814	AI306814 qw44g09.x
	32	174.8	9.2	227	44	AI308044	AI308044 tb16h08.x
	33	174.8	9.2	242	44	AI308888	AI308888 tb50b07.x
	34	174.8	9.2	242	44	AI309895	AI309895 tb48h06.x
	35	174.8	9.2	232	44	AI310859	AI310859 ta43c12.x
	36	174.8	9.2	228	44	AI311359	AI311359 ta47g08.x
	37	174.8	9.2	219	44	AI311967	AI311967 ta84d11.x
	38	174.8	9.2	217	45	AI340387	AI340387 tb40b10.x
	39	174.8	9.2	217	45	AI340423	AI340423 tb40f02.x
	40	174.8	9.2	217	47	AI473160	AI473160 th54g06.x
	41	174.2	9.2	234	44	AI305750	AI305750 qw44a11.x
	42	173.8	9.2	222	43	AI223570	AI223570 qx30g06.x
	43	173.8	9.2	218	43	AI223720	AI223720 qx32a03.x
	44	173.8	9.2	237	43	AI224320	AI224320 qw99a10.x
	45	173.8	9.2	220	43	AI224659	AI224659 qw97h03.x

ALIGNMENTS

```
RESULT   1
AA662164/c
LOCUS       AA662164      606 bp    mRNA            EST       03-DEC-1997
DEFINITION  ns66c07.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1188588 3'
            similar to SW:ETS4_DROME P29775 DNA-BINDING PROTEIN D-ETS-4 ;, mRNA
            sequence.
ACCESSION   AA662164
NID         g2616255
VERSION     AA662164.1  GI:2616255
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1392902.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 734   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 393.
FEATURES             Location/Qualifiers
     source          1..606
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1188588"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Prim. prostate, Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I  and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      104 a    172 c    204 g    126 t
ORIGIN
```

```
  Query Match             29.4%;  Score 556;  DB 36;  Length 606;
  Best Local Similarity   97.9%;  Pred. No. 7.6e-109;
  Matches  595;  Conservative    0;  Mismatches   10;  Indels    3;  Gaps    3;

Qy    1308 ttccatccgcaaaaaccatccccgccatgaactacgacaagctgagccggctccatccggcca 1347
           ||| |||||||||| ||||||||||||||||||||||||||||| |||||||| ||||
Db     606 GGGGATCCGCAAGAA-GGTCCCGGCCATGAACTACGACAAGCTGAGCGGCTCCATCGGGCA 548

Qy    1348 gtattacaaaaaaaaaatatcccggaagccagacatctcccagcgcctcgtctaccagtt 1407
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     547 GTATTACAAGAAGGGCATCATCCGGAAGCCAGACATTTCCCAGCGCCTCGTCTACCAGTT 488
```

```
Qy    1408 cgtgcaccccatctgagtgcctggcccagggcctgaaaccccgccctcaggggcctctctc 1467
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     487 CGTGCACCCCATCTGAGTGCCTGGCCCAGGGCCTGAAACCCCGCCTCAGGGGCCTCTCTC 428

Qy    1468 ctgcctgccctgcctcagccaggccctgagatgggggaaaacgagcaatctactctgctg 1527
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     427 CTGCCTGCCCTGCCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTCTACTCTGCTG 368

Qy    1528 ctctgaccttccagagcccaaggtcagggagggcaaccaactgccccagggggatatgg 1587
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     367 CTCTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGATATGG 308

Qy    1588 gtcctctgggggcttcgggaccatgagggcagggtgcttcctcctcaggcccagctgctc 1647
           |||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||
Db     307 GTCCTCTGGGG-CTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTC 249

Qy    1648 cctgggaggacagagggagacagggctgctccccaacacctgcctctgaccccagcattt 1707
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 CCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTT 189

Qy    1708 ccagagcagagcctacagaagggcagtgactcgacaaagccgcacagccagtccagccctc 1767
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db     188 CCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAGAGCCACAGCCAGTCCAGCCTC 129

Qy    1768 tctctgctccatccccctgcctcccattctgcaccacacctggcatggtgcaaggagaca 1827
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 TCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACA 69

Qy    1828 tctgcacccctgagttgggcagccaggaqtgcccccg-qqaatqqataataaaqatacta 1886
           |||||||||||||||| ||||||| |||||||||| ||||||||||||||||||||||
Db      68 TCTGCACCCCTGAGTTGTGCAGCCATGAGTGCCCCCGTGGAATGGATAATAAAGATACTA 9

Qy    1887 gagaactg 1894
           ||||||||
Db       8 GAGAACTG 1
```

```
RESULT   2
AA662204
LOCUS       AA662204      514 bp    mRNA            EST       03-DEC-1997
DEFINITION  ns66c07.r1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1188588 5'
            similar to SW:ETS4_DROME P29775 DNA-BINDING PROTEIN D-ETS-4 ;, mRNA
            sequence.
ACCESSION   AA662204
NID         g2616295
VERSION     AA662204.1  GI:2616295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1392997.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 734   Std Error: 0.00
            Seq primer: -28m13 revl ET from Amersham.
FEATURES             Location/Qualifiers
```

```
     source          1. .514
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1188588"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I  and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      148 a    173 c    151 g     82 t
ORIGIN
```

```
  Query Match             26.9%;  Score 509.2;  DB 36;  Length 514;
  Best Local Similarity   99.4%;  Pred. No. 6.1e-99;
  Matches  511;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    124 gaaggcgcatcttcaaaattgaggactcagcccaggtggcccggctgtggggcatccgcaa 1299
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAGGGCATCTTCAAAATTGAGGACTCAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAA 60

Qy    1300 gaaccatcccgccatgaactacgacaagctgagccgctccatccgccagtattacaagaa 1359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAACCATCCCGCCATGAACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAA 120

Qy    1360 gggcatcatccgaaagccagacatctcccagcgcctcgtctaccagttcgtgcaccccat 1419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGGCATCATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACCCCAT 180

Qy    1420 ctgagtgcctgggcccagggcctgaaacccgcccctcaggggcctctctcctgcctgccctg 1479
          |||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||
Db     181 CTGAGTGCCTGGACCAGGGCATGAAACCCGCCCTCAGGGGCCTCTCTCCTGCCTGCCCTG 240

Qy    1480 gctgctctgaccttccagagcccaaggtcagggagggggcaaccaactgccccagggggat 1539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCTGCTCTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGAT 300

Qy    1540 agaggcccaaggtcagggagggggcaaccaactgccccagggggatatgggtcctctggggc 1599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGAGGCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGC 360

Qy    1600 attcgaaccatgggcaggggtgcttcctcctcaggcccagctgctccccctggaggaca 1659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTTCGGGACCATGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACA 420

Qy    1660 tagggagacagggctgctccccaacacctgcctctgaccccagcatttccagagcagagc 1719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGC 480

Qy    1720 ctacaaaaggcagtgactcgacaaaggccacag 1753
          |||||||||||||||||||||||||||||||||
Db     481 CTACAAAAGGGCAGTGACTCGACAAAGGCCACAG 514


RESULT 1
AI668916
LOCUS       AI668916      512 bp    mRNA            EST       14-MAY-1999
DEFINITION  wb02802 x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2313507 3',
            mRNA sequence.
ACCESSION   AI668916
NID         g4833690
VERSION     AI668916.1  GI:4833690
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948749.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M  Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2313507"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986758, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       96 a    142 c    155 g    119 t
ORIGIN
```

```
  Query Match             24.6%;  Score 465.4;  DB 50;  Length 512;
  Best Local Similarity   96.7%;  Pred. No. 1.2e-89;
  Matches  475;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    1404 agttcgtgcaccccatctgagtgcctggcccagggcctgaaacccgcccctcagggggctc 1463
          |||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     512 AGTTTGTGCACCCCATTTGAGTGCCTGGCCCAGGGCCTGAAACCCGCCCTCAGGGGCTC 453

Qy    1464 tctcctgcctgccctgcctcaggcaggccctgagatgggggaaaacgggcagtctgctct 1523
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db     452 TTTCCTGCCTGCCCTGCCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTTTGCTTT 393

Qy    1524 gctgctctgaccttccagagcccaaggtcagggagggggcaaccaactgccccagggggat 1583
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 GCTGCTTTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGAT 333

Qy    1584 atgggtcctctggggccttcgggaccatgggcaggggtgcttcctcctcaggcccagct 1643
          ||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||| ||
Db     332 ATGGGTCCTTTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAACT 273

Qy    1644 gctccccctggaggacagagggagacagggctgctccccaacacctgcctctgaccccagc 1703
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     272 GCTCCCCTGGAGGACAAAGGGAGACAGGGCTGCTCCCCAACACCTGCCTTTGACCCCAGC 213

Qy    1704 atttccagagcagagcctacagaaaggcagtgactcgacaaaggccacaggcagtccaag 1763
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212 ATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGG 153

Qy    1764 cctctctctgctccatccccctgctcccattctgcaccacacctggcatggtccagggg 1823
```

```
          ||| | | ||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    152 CCTTTTTTTGCTCCATCCCCCTGCCTCCCATTTTGCACCACACCTGGCATGGTGCAGGGA 93

Qy   1424 gtaatctcagcagcctgaatttgaacagccaagaagtgccccctgggaatggataataaagata 1883
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 GACATTTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATA 33

Qy   1884 ctaaagaactg 1894
          ||||||||||||
Db     32 CTAGAGAACTG 22


RESULT   4
AA649821/c
LOCUS       AA649821      492 bp    mRNA            EST       23-OCT-1997
DEFINITION  nq84g11.s1 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1159076 3',
            mRNA sequence.
ACCESSION   AA649821
NID         g2563600
VERSION     AA649821.1  GI:2563600
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 392)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Nov 24, 1997 this sequence version replaced gi:430004.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .392
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1159076"
                     /clone_lib="NCI_CGAP_Co9"
                     /tissue_type="colon tumor RER+"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     RER+ colon tumor, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is not normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo
                     (Soares4). "
BASE COUNT       65 a    113 c    131 g     83 t
ORIGIN


  Query Match             19.8%;  Score 374.8;  DB 36;  Length 392;
  Best Local Similarity   99.5%;  Pred. No. 1.8e-70;
  Matches  376;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   1517 ctactctgctgctctgaactttccaagcccaaggtcagggagggggcaaccaaactgccccca 1576
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     60 CTGGTCTGGTGCTTGACCTTCCAGAGCCCAAGGTCAGGGAGGGCCAACCAACTGCCCCA 319
```

```
Qy   1577 ggggatatgggtcctctgggggccttcgggaccatgggcaggggtgcttcctcctcaaq 1636
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    318 GGGGGATATGGGTCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAAG 259

Qy   1637 cccagctgctccccctggaggacagagggagacagggctgctccccaacacctgcctctga 1696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    258 CCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGA 199

Qy   1697 ccccagcatttccagagcagagcctacagaagggcagtgactcgacaaaggccacaggca 1756
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    198 CCCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCA 139

Qy   1757 gtccaggcctctctctgctccatccccctgcctcccattctgcaccacaccctagcatggt 1816
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    138 GTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCCTGGCATGGT 79

Qy   1817 gcagggagacatctgcacccctgagttgggcagccaggagtgccccccgggaatggataat 1876
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     78 GCAGGGAGACATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAAT 19

Qy   1877 aaagatactagagaactg 1894
          ||||||||||||||||||
Db     18 AAAGATACTAGAGAACTG 1


RESULT   5
AI685592
LOCUS       AI685592      386 bp    mRNA            EST       27-MAY-1999
DEFINITION  tu20c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PTR5.t3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
NID         g4896886
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188477.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
```

circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 72 a 140 c 110 g 64 t
ORIGIN

Query Match 18.4%; Score 348.8; DB 50; Length 386;
Best Local Similarity 99.4%; Pred. No. 6e-65;
Matches 350; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 293 ctgaaacacgccagtggcctcagctgcccacacctcttcccggcccctgaagttggcact 349
Db 35 CTGGGACACGCCAGTGGCTCCAGCTGCCCACACCTCTTCCCGGCCCCTGAAGTTGGCACT 94

Qy 350 gcagcagacagctccctgggcaccaggcagctaacagacacagccgccagcccaaacagc 409
Db 95 TAGCAGACAGCTCCCTGGGCACCAGGCAGCTAACAGACACAGCCGCCAGCCCAAACAGC 154

Qy 410 agcagatgggcagctccagccccgggtctgagcagcgtatccccccagccacctcctgctg 469
Db 155 AGCCGCATGGGCAGCTCCAGCCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCTGCTG 214

Qy 470 cccccgtacacctatgtcccgaacaggcttggagaaggcggcagcgggggcagtgggtctc 529
Db 215 CCCCCGTACACCTATGTCCCGAACAGGCTTGGAGAAGGCGGCAGCGGGGGCAGTGGGTCTC 274

Qy 530 agaagacgggactgaatcccaagtccacccgccacgcccgagcagggcctgtccgccttc 589
Db 275 AGAGACGGGACTGAATCCCAGTCCACCCGCCACGCCCGAGCAGGGCCTGTCCGCCTTC 334

Qy 590 tacctctcctactttgacatgctgtaccctgaggacagcagctgggcagcca 641
Db 335 TACCTCTCCTACTTTGACATGCTGTACCCTGAGGACAGCAGCTGGGCAGCCA 386

RESULT 6
AA662521/c
LOCUS AA662521 366 bp mRNA EST 12-NOV-1997
DEFINITION nm96e07.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1218852 3', mRNA sequence.
ACCESSION AA662521
NID g2616612
VERSION AA662521.1 GI:2616612
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE 1 (bases 1 to 366)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 19, 1997 this sequence version replaced gi:1520374.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Seq primer: 40m13 fwd. ET from Amersham.
FEATURES Location/Qualifiers

source 1..366
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1218852"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 56 a 94 c 110 g 106 t
ORIGIN

Query Match 17.2%; Score 325.2; DB 36; Length 366;
Best Local Similarity 99.1%; Pred. No. 6e-60;
Matches 327; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1562 caaccaactgccccaggggcgatatgggtcctctgggggccttcgggaccatgggcaggggg 1621
Db 366 CAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGCTTTCGGGACCCTGGGGCAGGGG 307

Qy 1622 tgcttcctcctcaggcccagctgctcccctggaaggacagagggagacagaggctgctcccc 1681
Db 306 TGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCC 247

Qy 1682 aacacctgcctctgaccccagcatttccagagcagagcctacagaagggcaatgaactcga 1741
Db 246 AACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGAACTCGA 187

Qy 1742 caaaggccacaggcagtccaggcctctctctgctccatccccctgcctcccattctgcac 1801
Db 186 CAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTCCCTTTCTGCAC 127

Qy 1802 cacacctggcatggtgcagggagacatctgcacccctgagttgggcagccaggagtgccc 1861
Db 126 CACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCC 67

Qy 1862 ccgggaatggataataaagatactagagaa 1891
Db 66 CCGGGAATGGATAATAAAGATACTAGAGAA 37

RESULT 7
AI687669/c
LOCUS AI687669 366 bp mRNA EST 27-MAY-1999
DEFINITION tp82g09.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2205856 3', mRNA sequence.
ACCESSION AI687669
NID g4898963
VERSION AI687669.1 GI:4898963
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Feb 18, 1999 this sequence version replaced gi:4297630.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

```
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 365.
FEATURES              Location/Qualifiers
     source           1..366
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="11 q11-q13.4"
                      /clone="IMAGE:2205856"
                      /clone_lib="NCI_CGAP_Ut3"
                      /tissue_type="poorly-differentiated endometrial
                      adenocarcinoma, 2 pooled tumors"
                      /lab_host="DH10B"
                      /note="Vendor: uterus, Vector: pCMV-SPORT6, Site_1: SalI,
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.45 kb. Life Technologies catalog #:
                      11541-018"
BASE COUNT        58 a    104 c    113 g     90 t      1 others
ORIGIN


  Query Match             17.1%;  Score 323.6;  DB 50;  Length 366;
  Best Local Similarity   98.3%;  Pred. No. 1.3e-59;
  Matches  348;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy    1542 ctcccaaggtcaaggaaggggcaaccaactgccc-caggggggatatgggtcctctgggccc 1600
           ||||||| ||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db     466 AGCCCAAGGTCAGGGAGGGGCAACCAACTG-CCNCAGGGGGATATGGGTCCTCTGGGGCC 308

Qy    1601 tttcaggaccatgaggcaggatgcttcctcctcaggcccagctgctcccctgaaggacag 1660
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     307 TTCGGGACCCTGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGAAGGACAG 248

Qy    1661 agggaagcaggactgctcccaacacctgcctctgacccagcatttccagagcagagcc 1720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247 AGGCAGACAGGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCC 188

Qy    1721 tacaggaaggccagtgactcaacaaaggccacaggcagtccaggcctctctctgctccatc 1780
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     187 TACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTTTGCTCCATC 128

Qy    1781 ccccagcctccccattctgcaccacacctggcatggtgcagggagacatctgcaccccctga 1840
           ||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     127 CCCCTGCCTCCCATTTTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGA 68

Qy    1841 attgggcaaccaaaagtgccccccaggaatggataataaagatactagagaaactg 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67 TTCCGGCAGCCACAGTGCCCCCGGAATGGATAATAAAGATACTAGAGAACTG 14


RESULT   8
AA513274
LOCUS       AA513274      455 bp    mRNA            EST       19-AUG-1997
DEFINITION  nq04e12.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:942474, mRNA
            sequence.
ACCESSION   AA513274
NID         g2251686
VERSION     AA513274.1  GI:2251686
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I M A G E  Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 588   Std Error: 0.00
            Seq primer: -40ml3 fwd, ET from Amersham
            High quality sequence stop: 319.
FEATURES              Location/Qualifiers
     source           1   355
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:942474"
                      /clone_lib="NCI_CGAP_Pr6"
                      /sex="male"
                      /tissue_type="prostate"
                      /lab_host="DH10B"
                      /note="Vector: pAMP10; mRNA made from prostatic
                      intraepithelial neoplasia (low-grade), cDNA  made by
                      oligo-dT priming.  Non-directionally cloned.
                      Size-selected  on agarose gel, average insert size 600 bp.
                      Reference:  Krizman et al. (1996) Cancer Research
                      56:5380-5383."
BASE COUNT        67 a    121 c    109 g     58 t
ORIGIN


  Query Match             16.8%;  Score 317.6;  DB 34;  Length 355;
  Best Local Similarity   98.3%;  Pred. No. 2.4e-58;
  Matches  342;  Conservative    0;  Mismatches    4;  Indels    2;  Gaps    2;

Qy    1430 ggcccagggcctgaaaacccgccctcagggggctctctcctgcctgccctgcctcagccag 1489
           ||| |||||| |||||||||||||||| || ||||||||||||||||||||||||||||||
Db       9 GGCACAGGGCATGAAACCCGCCCTCAGTGG-CTCTCTCCTGCCTGCCCTGCCTCAGCCAG 67

Qy    1490 gccctgagatgggggaaaacgggcagtctgctctgctgctctgaccttccagaggcccaag 1549
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 GCCCTGAGATGGGGGAAAACGGGCAGTCTGCTCTGCTGCTCTGACCTTCCAGAGGCCAAG 127

Qy    1550 gtcagggagggggcaaccaactgccccagggggatatgggtcctctgggggcttcaggacc 1609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 GTCAGGGAGGGGCAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGCCTTCGGGACC 187

Qy    1610 atgggcagggggtgcttcctcctcaggcccagctgctcccctgaggacagaggagaaca 1669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188 CTGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGAAGGACAGAGGAGACA 247

Qy    1670 gggctgctccccaacacctgcctctgacccagcatttccagagcagagcctacagaagg 1729
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 GGGCTGCTCCCCAACACCTGCCTCTGACCCAGCATTTCCAGAGCAGAGCCTACAGAAGG 307

Qy    1730 gcagtgactcgacaaaggccacaggcag-tccaggcctctctctgctc 1776
           |||||||||||||||||||||||||||| ||||||||||||||||||
Db     308 GCAGTGACTCGACAAAGGCCACAGGCAGTTCCAGGCCTCTCTCTGCTC 355


RESULT   9
AA548598/c
LOCUS       AA548598      384 bp    mRNA            EST       14-AUG-1997
DEFINITION  nk54e12.s1 NCI_CGAP_Pr7 Homo sapiens cDNA clone IMAGE:1017334, mRNA
            sequence.
ACCESSION   AA548598
NID         g2318880
```

```
VERSION     AA548598.1  GI:2318880
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400848

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..384
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1017334"
                     /clone_lib="NCI_CGAP_Pr7"
                     /sex="male"
                     /tissue_type="low-grade prostatic neoplasia"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pAMP10; mRNA made from
                     prostate intraepithelial neoplasia (high-grade), cDNA
                     made by oligo-dT priming.  Non-directionally cloned.
                     Size-selected  on agarose gel, average insert size 600 bp
                     "
BASE COUNT       93 a    111 c    118 g     92 t
ORIGIN
```

```
  Query Match             16.3%;  Score 308.4;  DB 35;  Length 384;
  Best Local Similarity   99.4%;  Pred. No. 2.2e-56;
  Matches  320;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy    1573 ccccagggggatatgggtcctctgggggccttcgggaccatggggcaggggtgcttcctcct 1632
           |||||||||||||||.|||||||||||| |||||||||| ||||||||||||||||||||||
Db     324 CCCAGGGGGATATGGGTCCTCTGGGG-CTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCT 326

Qy    1633 cagaaccaactgctcccctggaggacagagggagacagggctgctccccaacacctgcct 1692
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325 CAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCT 266

Qy    1693 ctgagcccagcatttccagagcagagcctacagaagggcagtgactcgacaaaggccaca 1752
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     265 CTGACCCCAGCAATTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACA 206

Qy    1753 tgcagtccaaatcctctctgctgctccatcccctgcctcccattctgcaccacaactggca 1812
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 GGCAGTCCAGGCCTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACAACTGGCA 146

Qy    1813 tgatgcagggagacatctgcaccctgagttgggcagccaggagtgcccccgggaatgga 1872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 TGATGCAGGGAGACATCTGCACCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGA 86

Qy    1873 taataaagatactagaggaactg 1894
           |||||||||||||||||||||||
Db      85 TAATAAAGATACTAGAGGAACTG 64
```

```
RESULT  10
AA228304
LOCUS       AA228304      369 bp    mRNA            EST       21-AUG-1997
DEFINITION  nc38e07.r1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1010436, mRNA
            sequence.
ACCESSION   AA228304
NID         g1849875
VERSION     AA228304.1  GI:1849875
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393445.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 731   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1..369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1010436"
                     /clone_lib="NCI_CGAP_Pr2"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from
                     5,000-10,000 microdissected preneoplastic cells
                     histologically-determined to be prostatic intraepithelial
                     neoplasia 2 (PIN2) cells.  Double-stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG-cloning
                     method (Life Technologies).  Average insert size is 600
                     bp. NOTE: Not directionally cloned.  This library was
                     constructed by  David Krizman."
BASE COUNT       70 a    122 c     98 g     79 t
ORIGIN
```

```
  Query Match             15.4%;  Score 291.2;  DB 30;  Length 369;
  Best Local Similarity   99.0%;  Pred. No. 9.7e-53;
  Matches  293;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    1458 ggcctctctcctgcctgccctgcctcagccaggccctgaagatgggggaaaacgaggcagtc 1517
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74 GGCCTCTCTCCTGCCTGCCCTGCCTCAGCCAGGCCCTGAGATGGGGGAAAACGAGGCAGTC 133

Qy    1518 tgctctgctgctctgaccttccagagcccaaggtcagggaggggcaaccaactgccccag 1577
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     134 TGCTCTGCTGCTCTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAG 193

Qy    1578 ggggatatgggtcctctgggggccttcgggaccatggggcaggggtgcttcctcctcaggc 1637
```

```
          ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    194 GGGACTATGGGTCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGC 253

Qy   1638 ccagcctgctcccctgaaaaacagaggaagacagggctgctccccaacacctgcctctgac 1697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 CCAGCTGCTGCCCTGGAGCACAGAGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGAC 313

Qy   1698 cccagcatttccagagcagagcctacagaagggcagtgactcgacaaaggccacag 1753
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 CCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAG 369


RESULT 11
AA940471
LOCUS       AA940471      504 bp    mRNA            EST       23-APR-1998
DEFINITION  zo45a05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone IMAGE:1052816 5' similar to SW:ETS4_DROME P29775 DNA-BINDING
            PROTEIN D-ETS-4 ;, mRNA sequence.
ACCESSION   AA940471
NID         g3079964
VERSION     AA940471.1  GI:3079964
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2287074.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St  Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL , contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:584492
            Seq primer: -28ml4 rev2 ET from Amersham
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1..504
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1052816"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'IGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG],  digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead.  "
BASE COUNT      122 a    149 c    135 g     98 t
ORIGIN


  Query Match          13.7%;  Score 259.8;  DB 40;  Length 504;
  Best Local Similarity   82.5%;  Pred. No. 4.9e-46;
  Matches  311;  Conservative    0;  Mismatches   62;  Indels    4;  Gaps    1;

Qy   1109 gaggagagctggaccgacagcgaggtggactcatcatgctccggacagcccatccacctg 1168
          |||||| ||||||| ||  | |||||||||| || ||||||||||||||| ||||||
Db      1 GAGGAGGGCTGGACGGATGGTGAGGTGGACTCGTCGTGCTCCGGGCAGCCCATTCACCTG 60

Qy   1169 tggcagttcctcaaggagttgctactcaagccccacagctatggccgcttcattaaqtaq 1228
          |||||||||||| || ||  |||| ||||||||||||||||||||||||||||  | |||
Db     61 TGGCAGTTCCTGAAAGAACTGCTGCTCAAGCCCCACAGCTATGGCCGCTTCATCGGCTGG 120

Qy   1229 ctcaacaaggagaaagggcatcttcaaaattgaggactcagcccaggtggcccggctgtgg 1288
          ||||||||||||||| ||||||||||||||||||||||||||| |||||||||| ||||||
Db    121 CTCAACAAGGAGAAAGGCATCTTCAAAATTGAGGACTCAGCACAGGTGGCCCGACTGTGG 180

Qy   1289 ggcatccgcaagaaccgtcccgccatgaactacgacaagctgaggccgctccatccgccag 1348
          ||  | |||||||||||| || |||||||||| || || || ||||||||||||||||||
Db    181 GGTGTGCGCAAGAACCGGCCAGCCATGAACTATGATAAACTAAGCCGCTCCATCCGCCAG 240

Qy   1349 tattacaagaagggcatcatccggaagccagacatctcccagcggcctcgtctaccagttc 1408
          |||||||||||||||||||| || || || |||||||| |||||||| |||||||| ||
Db    241 TATTACAAGAAGGGCATCATTCGTAAACCCGACATCTCTCAGCGCCTTGTCTACCAATTT 300

Qy   1409 gtgcaccccatctgagtgcc----tggcccagggcctgaaaccccgccctcaqqqqcctct 1464
          |||| || |||||| |||     | ||   ||||| |||| ||||    | ||| ||
Db    301 GTGCATCCAGTCTGAGAGCCACAGAGACCAGAGGCCTACAACCTGCCCCAGGCAGCCACT 360

Qy   1465 ctcctgcctgccctgcc 1481
          ||| |   |||  | |
Db    361 CTCTGGTTGGCCTGGTC 377


RESULT 12
AI686414/c
LOCUS       AI686414      441 bp    mRNA            EST       27-MAY-1999
DEFINITION  tu41g05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253656 3',
            mRNA sequence.
ACCESSION   AI686414
NID         g4897708
VERSION     AI686414.1  GI:4897708
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189306.

            Contact: Robert Strausberg, Ph D
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M D , Ph D , Michael R
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..441
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2253656"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
```

/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-988759, 1181192-1181959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 46 a 94 c 104 g 143 t 4 others
ORIGIN

Query Match 12.5%; Score 236.2; DB 50; Length 441;
Best Local Similarity 82.0%; Pred. No. 4.8e-41;
Matches 236; Conservative 0; Mismatches 63; Indels 2; Gaps 2;

Qy 1544 [illegible] 1593
Db 439 [illegible] 349

Qy 1594 [illegible] 1653
Db 348 [illegible] 289

Qy 1654 [illegible] 1713
Db 288 [illegible] 229

Qy 1714 caaaaccctacaaaaaggcagtgactcga-caaaggccacaggcagtccaggcctctctct 1772
Db 228 AAAGCCTTCAAAAAGGCAGTGTCT-TGNCAAAGGCCCCAGGCAATTCAGGCCTTTTTTT 170

Qy 1773 [illegible] 1832
Db 169 GGTCCATCCCCCTGCCTCCCATTTTGCCCCACACCTGGGATGGGGCAAGGAAAAATTTTC 110

Qy 1833 acccctgaattgtggcaaccaggagtgccccccgggaatggataataaagatactagagaac 1892
Db 109 TCCCCTTAGTTGGGCAACCAGGAGGGCCCCCGGGAATGGATAATAAAAATTCTAAAGAAC 50

Qy 1893 * 1893
Db 49 1 49

RESULT 79
AI451780
LOCUS AI451780 544 bp mRNA EST 09-MAR-1999
DEFINITION mk22i07.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone IMAGE:330181 5', mRNA sequence.
ACCESSION AI451780
NID g4306393
VERSION AI451780.1 GI:4306393
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 544)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948093.

Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1810
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the correct orientation)
Seq primer: -40PP from Gibco
High quality sequence stop: 462.

FEATURES Location/Qualifiers
source 1..544
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:330181"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

BASE COUNT 104 a 190 c 153 g 97 t
ORIGIN

Query Match 10.9%; Score 207; DB 46; Length 544;
Best Local Similarity 66.4%; Pred. No. 8e-35;
Matches 367; Conservative 0; Mismatches 165; Indels 21; Gaps 4;

Qy 80 cagctccctgcaagccccctcaggttgggccttgccacggtgccagcagggcagccctgggc 139
Db 8 CTGCCCCGGTGAGAACCCCCAGCTTGGGCCTTGTCATGGTGCCAGCAGGTGGCCCTGAGC 67

Qy 140 tgggggtagggactccctacaggcacgcagccctgagacctcagaagggccaccccttga 199
Db 68 TTCTGACAGGGGCCTGCCTATAGACCTGCAGGCCTGAGGCCICAGA----CICACACICA 123

Qy 200 gggtggccaggcccccagtggccaacctgagtgctgcctctgccaccagccctgctgggcc 259
Db 124 AGGGGCAAGAGGCCCTGGTGGCCCAGCTAAGAGCCACCTCTGTCCCCAGCCCTGCTGGCC 183

Qy 260 cctggttccgctggccccccagatgcctggctgagacacgccagtgggctcagctgccca 319
Db 184 CCACTGATGTCTGAC----------TGAGACCCAGCAGTGACCCTGAGCTGGCTGCCCAC 233

Qy 320 cacctcttcccggcccctgaagttggcactgcag····cagacagctccctgaggcaccag 375
Db 234 TGCCTCCTCCTGGTCCCTGAGGTTGGCTCTGCCGAGGACGGACGACICTICTGACAGCAG 293

Qy 376 gcagctaacagacacagccgccagcccaaacagcagcggcatgggcagcgccagccccgg 435
Db 294 GCGGCTAACGGAAGCAGCCCCAAGCCTCCACCGCA---GCATGGGCAGTGCCAGCCCAGG 350

Qy 436 tctgagcagcgtatcccccagccacctcctgctgccccccgacacggtgtcggcggacagg 495
Db 351 CCTGAGCAACGTGTCCCCCGGTTGCCTGCTACTGTTCCCAGATGTGGCACCACGAACAGG 410

Qy 496 cttggagaaaggcggcagcgggggcagtgggtctcgagagacgggactgagtccgactcc 555
Db 411 GACGGAGAAGGCAGCATCAGGAGCAAIGGGCCCIGAGAAGCAGGAATGGAGTCCIAGICC 470

Qy 556 acccgccacgcccgagcagggcctgtccgccttctaccctctcctactttgacatgctgta 615
Db 471 ACCCGCCACCCCTGAGCAGGGCCIGICTGCTTTCTACCICTCTTACTTTAACATGTATCC 530

Qy 616 ccctgaggacagc 628

```
                |   ||   ||||
Db     541 CGACCATAGCAGC 543


RESULT  14
R75783
LOCUS       R75783          390 bp    mRNA            EST       06-JUN-1995
DEFINITION  yz12e12.r1 Soares breast 2NbHBst Homo sapiens cDNA clone
            IMAGE:196402 5', mRNA sequence.
ACCESSION   R75783
NID         g850465
VERSION     R75783.1  GI:850465
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 568
            High quality sequence stops: 329
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 568   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 329
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Homo sapiens"
                     /db_xref="GDB:572958"
                     /db_xref="taxon:9606"
                     /map="10"
                     /clone="IMAGE:158902"
                     /clone_lib="Soares breast 2NbHBst"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     IGIIAGCAAIGIGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                     Library went through one round of normalization to a Cot -
                     23%. Library constructed by Bento Soares and M Fatima
                     Bonaldo."
BASE COUNT       78 a    118 c    112 g     82 t
ORIGIN


  Query Match             9.5%;  Score 180.2;  DB 22;  Length 390;
  Best Local Similarity   94.7%;  Pred. No. 3.5e-29;
  Matches  230;  Conservative    0;  Mismatches    8;  Indels    5;  Gaps    4;

Qy    151 gcaatctgctctgctgctctgacctccagaggcccaaggtcagggaggggcaaccaactg 1571
          | |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 GAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTG 188

Qy   1572 ccccagggggatatgggtcctctgggggccttcgggaccatgggacagggtgcttcctcc 1631
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    189 CCCCAGGGGGATATGGGTCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCC 248

Qy   1632 tcaggcccagctgctccctgg--aggacagagggagacagggctgctccccaacacctg 1689
          ||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db    249 TCAGGCCCAGCTGCTCCCCTGGGAGGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTG 308

Qy   1690 cctc-tgaccccagca-tttccagaagcagagcc-tacagaaggggcagtgactcgacaaag 1746
          |||| |||||||||| |||||||||||||||| |||||||||||||| |||||||||
Db    309 CCTCTTGACCCCAGCATTTTCCAGAGCAGAGCCTTACAGAAGGGCAGTGATTCGACAAAG 368

Qy   1747 gcc 1749
          | ||
Db    369 GGC 371


RESULT  15
AI305722
LOCUS       AI305722      245 bp    mRNA            EST       03-FEB-1999
DEFINITION  qw47g09.x1 NCI_CGAP_Br13 Homo sapiens cDNA clone IMAGE:1994272 3',
            mRNA sequence.
ACCESSION   AI305722
NID         g3990613
VERSION     AI305722.1  GI:3990613
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 245)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 330   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .245
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1994272"
                     /clone_lib="NCI_CGAP_Br13"
                     /sex="female"
                     /tissue_type="breast carcinoma in situ"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     carcinoma, cDNA made by oligo-dT  priming.  Directionally
                     cloned.  Size-selected on agarose  gel, average insert
                     size 500 bp.  Primary library, non-amplified.  "
BASE COUNT       83 a     66 c     56 g     40 t
ORIGIN


  Query Match             9.3%;  Score 175.2;  DB 44;  Length 245;
  Best Local Similarity   95.7%;  Pred. No. 3.5e-28;
  Matches  180;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy   1707 tccagaagcagagcctacagaaggggcagtgactcgacaaagggccacagccagtccagggcct 1766
          | | | |   || | ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      18 TGCTGCGAGAAGACGACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCT 77

Qy    1767 ctctctgctccatccccctgcctcccattctgcaccacacctggcatggtgcaqggagac 1826

Db      78 CTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGAC 137

Qy    1827 atctgcacccctgagttgggcagccaggagtgccccccgggaatggataataaagatacta 1886

Db     138 ATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTA 197

Qy    1887 gaqaactg 1894

Db     198 GAGAACTG 205


Search completed: November 23, 1999, 15:57:13
Job time: 587 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 1999, 17:57:57 ; Search time 307.18 Seconds
(without alignments)
12162.180 Million cell updates/sec

Title: US-09-126-945-1
Perfect score: 1894
Sequence: 1 atctgacttgctcccagcac.....ataaagatactagagaaactg 1894

Scoring table: OLIGO_NUC

Searched: 2546578 seqs, 985266752 residues

Database: EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est9:*
28: gb_est4:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	410	21.6	514	36	AA662204	AA662204 ns66c07.r
2	349	18.4	386	50	AI685592	AI685592 tu20c07.x
c 3	284	15.0	384	35	AA548598	AA548598 nk54c12.s
c 4	284	15.0	392	36	AA639821	AA639821 nq84g11.s
5	247	13.0	355	34	AA513274	AA513274 nq94e10.s
c 6	230	12.1	366	36	AA662521	AA662521 nu99e07.s
7	173	9.1	216	43	AI223644	AI223644 qx31f05.x
8	173	9.1	223	44	AI270872	AI270872 qw67c10.x
9	173	9.1	222	44	AI271046	AI271046 qw51g09.x
10	173	9.1	244	44	AI275319	AI275319 qw66a10.x
11	173	9.1	232	44	AI275321	AI275321 qw66a12.x
12	173	9.1	224	44	AI284816	AI284816 qw25b04.x
13	173	9.1	199	44	AI289687	AI289687 qw42e08.x
14	173	9.1	220	44	AI289725	AI289725 qw42h11.x
15	173	9.1	218	44	AI305465	AI305465 qw46e08.x
16	173	9.1	218	44	AI305499	AI305499 qw46h09.x
17	173	9.1	245	44	AI305722	AI305722 qw47a09.x
18	173	9.1	231	44	AI305972	AI305972 qw48a01.x
19	173	9.1	231	44	AI305974	AI305974 qw48a03.x
20	173	9.1	222	44	AI305988	AI305988 qw48h09.x
21	173	9.1	212	44	AI306749	AI306749 qw44a10.x
22	173	9.1	220	44	AI306796	AI306796 qw44t02.x
23	173	9.1	215	44	AI306814	AI306814 qw44a09.x
24	173	9.1	227	44	AI308044	AI308044 tb16h08.x
25	173	9.1	242	44	AI308888	AI308888 tb50b07.x
26	173	9.1	242	44	AI309895	AI309895 tb48h06.x
27	173	9.1	232	44	AI310859	AI310859 ta43c12.x
28	173	9.1	228	44	AI311359	AI311359 ta47g08.x
29	173	9.1	219	44	AI311967	AI311967 ta84d11.x
30	173	9.1	217	45	AI340387	AI340387 tb40b10.x
31	173	9.1	217	45	AI340423	AI340423 tb40f02.x
32	173	9.1	217	47	AI473160	AI473160 th54g06.x
33	172	9.1	222	43	AI223570	AI223570 qx30g06.x
34	172	9.1	218	43	AI223720	AI223720 qx32a03.x
35	172	9.1	237	43	AI224320	AI224320 qw99a10.x
36	172	9.1	220	43	AI224659	AI224659 qw97h03.x
37	172	9.1	218	44	AI250521	AI250521 qx33b11.x
38	172	9.1	247	44	AI251236	AI251236 qw55g03.x
39	172	9.1	247	44	AI254601	AI254601 qw54e10.x
40	172	9.1	247	44	AI255044	AI255044 qw53g08.x
41	172	9.1	218	44	AI270807	AI270807 qw51c05.x
42	172	9.1	238	44	AI270910	AI270910 qw52a01.x
43	172	9.1	218	44	AI270922	AI270922 qw52b04.x
44	172	9.1	221	44	AI271043	AI271043 qw51g04.x
45	172	9.1	219	44	AI271066	AI271066 qw52g11.x

ALIGNMENTS

RESULT 1
AA662204
LOCUS AA662204 514 bp mRNA EST 03-DEC-1997
DEFINITION ns66c07 r1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1188588 5' similar to SW:ETS4_DROME P29775 DNA-BINDING PROTEIN D-ETS-4 ;, mRNA sequence.
ACCESSION AA662204
NID g2616295

```
VERSION     AA662204.1  GI:2616295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   [illegible] http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1392997.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 734  Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1188588"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I  and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      108 a    173 c    151 g     82 t
ORIGIN
```

```
  Query Match             21.6%;  Score 410;  DB 36;  Length 514;
  Best Local Similarity   99.6%;  Pred. No. 5.1e-197;
  Matches  510;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    1242 aaggcatcttcaaattgaagaactcaagcccaggtgacccggctgtggggcatccgcaaga 1301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7 AGGGCATCTTCAAAATTGAGGACTCAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAAGA 62

Qy    1302 aggcgccgcgcatgaactacgacgaactgtaggcgctccatccgccagtattacaagaaagg 1361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 AGCGCCGCGCATGAACTACGACGAACTGTAGGCGCTCCATCCGCCAGTATTACAAGAAGG 122

Qy    1362 gcatcatccgcgaaaccgacatcaccgacctcgtgctctaccagttcgtgcaccccatct 1421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 GCATCATCCGCGAAACCGACATCACCGACCTCGTGCTCTACCAGTTCGTGCACCCCATCT 182

Qy    1422 ctctgccctgaggggcctctctcctgcctgccctgcc 1481
           |||||||||| |||||| ||||||||||||||||||||
Db     183 CTCTGCCCTGAGGGGCCTCTCTCCTGCCTGCCCTGCC 242

Qy    1482 tcagccaggcctgagatgggggaaaacgggcagtctgctctgctgctctgaccttccag 1541
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     243 TCAGCCAGGCCTGAGATGGGGGAAAACGGGCAGTCTGCTCTGCTGCTCTGACCTTCCAG 302

Qy    1542 agccccaaagtcaaggaggggcaaccaactgccccagggggatatgggtcctctgggggcct 1601
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 AGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGCCT 362

Qy    1602 tcgggaccatggggcaggggtgcttcctcctcaggcccagctgctcccctgaaaacaaa 1661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     363 TCGGGACCATGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACAGA 422

Qy    1662 gtggagacagggctgctccccaacacctgcctctgaccccagcatttccagagcagagcct 1721
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423 GGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCCT 482

Qy    1722 acagaaagggcagtgactcgacaaaaggccacag 1753
           |||||||||||||||||||||||||||||||||
Db     483 ACAGAAGGGCAGTGACTCGACAAAGGCCACAG 514
```

```
RESULT   2
AI685592
LOCUS       AI685592      386 bp    mRNA            EST       27-MAY-1999
DEFINITION  tu20c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PTR5.t3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
NID         g4896886
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria, Primates; Catarrhini, Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:4188477.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       72 a    140 c    110 g     64 t
ORIGIN
```

```
  Query Match             18.4%;  Score 348;  DB 50;  Length 386;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-165;
  Matches  348;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     244 gacacgccagtgacctcagctgcccacacctcttcccggcccctgaagttggcactgcag 353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44 GACACGCCAGTGGGCTCAGCTGCCCACACCTCTTCCCGGCCCCTGAAGTTGGCACTGCAG 98

Qy     354 caaacaactcccctgggcaccaggcagctaacagacacagccgccagcccaaacagcagcg 413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99 CAGACAGTTCCCTGGGCACCAGGCAGCTAACAGACACAGCCGCCAGCCCAAACAGCAGCG 158

Qy     414 gcatgaaacagccaagccccaggtctgagcagcgtatcccccagccacctcctgctgcccc 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 GCATGGGCAGCGGCAGCCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCTGCTGCCCC 218

Qy     474 cgaacacatatgtcgcggacaggcttggagaaggcggcagcgggggcagtgggtctcgaga 533
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 CGAACACATGTCGCGGGACAGGCTTGGAGAAGGCGGCAGCGGGGGCAGTGGGTCTCGAGA 278

Qy     534 gacgggactgaagtcccagtccaccgccacacgccgagcagggcctgtccgccttctacc 593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279 GACGGGACTGAGTCCCAGTCCACCGCCACACGCCGAGCAGGGCCTGTCCGCCTTCTACC 338

Qy     594 tgtcctactttgacatgctgtaccctgaggacagcagctgggcagcca 641
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     339 TGTCCTACTTTGACATGCTGTACCCTGAGGACAGCAGCTGGGCAGCCA 386


RESULT   3
AA548598/c
LOCUS       AA548598      384 bp    mRNA            EST       14-AUG-1997
DEFINITION  nb74g12.s1 NCI_CGAP_Pr7 Homo sapiens cDNA clone IMAGE:1017334, mRNA
            sequence.
ACCESSION   AA548598
NID         g2318880
VERSION     AA548598.1  GI:2318880
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12  1996 this sequence version replaced gi:1400848

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..384
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1017334"
                     /clone_lib="NCI_CGAP_Pr7"
                     /sex="male"
                     /tissue_type="low-grade prostatic neoplasia"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pAMP10; mRNA made from
                     prostate intraepithelial neoplasia (high-grade), cDNA
                     made by oligo-dT priming.  Non-directionally cloned.
                     Size-selected on agarose gel, average insert size 600 bp
                     "
BASE COUNT       63 a    111 c    118 g     92 t
ORIGIN


  Query Match             15.0%;  Score 284;  DB 35;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 3e-133;
  Matches  284;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1611 tggggcaggggtgcttcctcctcaggcccagctgctcccctggaggacagaggggacacag 1670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     347 TGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGGACACAG 288

Qy    1671 ggctgctccccaacacctgcctctgaccccagcatttccagaagcagaggcctacagaaagg 1730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     287 GGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGGCCTACAGAAGGG 228

Qy    1731 cagtgactcgacaaaggccacaggcagtccaggcctctctctgctccatccccctgcctc 1790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     227 CAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTC 168

Qy    1791 ccattctgcaccacacctggcatggtgcaggagagacatctgcaccccctgagtttgggcagc 1850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     167 CCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAGTTGGGCAGC 108

Qy    1851 caggagtgccccccgggaatggataataaagatactagagaactg 1894
          |||||||||||||||||||||||||||||||||||||||||||||
Db     107 CAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 64


RESULT   4
AA639821/c
LOCUS       AA639821      392 bp    mRNA            EST       23-OCT-1997
DEFINITION  nq84g11.s1 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1159076 3',
            mRNA sequence.
ACCESSION   AA639821
NID         g2563600
VERSION     AA639821.1  GI:2563600
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 392)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:430004.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..392
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1159076"
                     /clone_lib="NCI_CGAP_Co9"
                     /tissue_type="colon tumor RER-"
                     /lab_host="DH10B"
```

/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from RER: colon tumor, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo (Soares4)."
BASE COUNT 65 a 113 c 131 g 83 t
ORIGIN

Query Match 15.0%; Score 284; DB 36; Length 392;
Best Local Similarity 100.0%; Pred. No. 3e-133;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1611 tagggcagggatgcttcctcctcaggcccagctgctccccctggaggacagagggagacag 1670
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 294 TGGGGCAGGGGTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAG 225

Qy 1671 gactgctccccctacacctgcctctgtaccccagcatttccagagcagagcctacagaagggg 1730
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 224 GGCTGCTCCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCCTACAGAAGGG 165

Qy 1731 cagtgactgctgacaaaaggccacaggcagtccaggcctctctctgctccatcccccctgcctc 1790
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 164 CAGTGACTGCTGACAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTC 105

Qy 1791 ccattctgtgcacacacctgtgcatgtgtgcaggcagacatctgcaccccctgagttgggcagc 1850
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104 CCATTCTGTGCACACACCTGTGCATGGTGCAGGGAGACATCTGCACCCCTGAGTTGGGCAGC 45

Qy 1851 cagaagtgccccccgcaaatgaataataaagatactagagaaactg 1894
||||||||||||||||||||||||||||||||||||||||||||
Db 44 CAGGAGTGCCCCCCGGAATGGATAATAAAGATACTAGAGAACTG 1

RESULT 5
AA513274
LOCUS AA513274 455 bp mRNA EST 19-AUG-1997
DEFINITION na94e10.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:942474, mRNA sequence.
ACCESSION AA513274
NID g2251686
VERSION AA513274.1 GI:2251686
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 455)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/btep/image/image.html

Insert Length: 588 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 319.
FEATURES Location/Qualifiers
source 1..355
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:942474"
/clone_lib="NCI_CGAP_Pr6"
/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from prostatic intraepithelial neoplasia (low-grade), cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT 67 a 121 c 109 g 58 t
ORIGIN

Query Match 13.0%; Score 247; DB 34; Length 355;
Best Local Similarity 99.7%; Pred. No. 1.6e-114;
Matches 297; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1461 ctctctcctgcctgccctgcctcagccaggccctgagatgggggaaaacgggcagtctgc 1520
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 39 CTCTCTCCTGCCTGCCCTGCCTCAGCCAGGCCCTGAGATGGGGGAAAACGGGCAGTCTGC 98

Qy 1521 tctgctgctctgaccttccagagcccaagqtcagggaggggcaaccaactgccccagggg 1580
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 99 TCTGCTGCTCTGACCTTCCAGAGCCCAAGGTCAGGGAGGGGCAACCAACTGCCCCAGGGG 158

Qy 1581 gatatgggtcctctggggccttcgggaccatggggcaggggtgcttcctcctcaggccca 1640
||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db 159 GATATGGGTCCTCTGGGGCCTTCGGGACCCTGGGGCAGGGGTGCTTCCTCCTCAGGCCCA 218

Qy 1641 gctgctcccctggaggacagagggagacagggctgctccccaacacctgcctctgaccccc 1700
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 219 GCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCCCAACACCTGCCTCTGACCCC 278

Qy 1701 agcatttccagagcagagcctacagaagggcagtgactcgacaaaggccacaggcagt 1758
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 279 AGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGT 336

RESULT 6
AA662521/c
LOCUS AA662521 366 bp mRNA EST 12-NOV-1997
DEFINITION nu93e07.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1218852 3', mRNA sequence.
ACCESSION AA662521
NID g2616612
VERSION AA662521.1 GI:2616612
KEYWORDS EST.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 19, 1997 this sequence version replaced gi:1520474.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham.

```
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1218852"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker, 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I  and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       56 a     94 c    110 g    106 t
ORIGIN
```

```
  Query Match             12.1%;  Score 230;  DB 36;  Length 366;
  Best Local Similarity   99.6%;  Pred. No. 6.4e-106;
  Matches  248;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1671 tagggcaaaggtgattttgtctctaaggcccagctgctccccctggaggacagagggagacag 1730
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     258 TAGGGCAAAGGTGATTTTGTCTCTAAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAG 199

Qy    1731 aactactccccaaaacctactctgaccccagcatttccagagcagagcctacagaaggg 1790
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     198 GGCTGCTCCCCAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCCTACAGAAGGG 139

Qy    1791 aattgactgcagcaaaggccacaggcagtccaggcctctctctgctccatccccctgcctc 1850
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     138 AATTGACTGCAACAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTC 79

Qy    1851 aattctgtgcaccacacctggcatggtgcagggagacatctgcaccctgagttgggcagc 1910
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78 AATTCTGTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCTGAGTTGGGCAGC 19

Qy    1911 aaaaatgccccctgggaatggataataaagatactagagaa 1951
           |||||||||||||||||||||||||||||||||||||||||
Db      18 AAAAATGCCCCCGGGAATGGATAATAAAGATACTAGAGAA 58
```

```
RESULT   7
AI223644
LOCUS       AI223644      216 bp    mRNA            EST       21-DEC-1998
DEFINITION  qw67c05.x1 NCI_CGAP_Ov34 Homo sapiens cDNA clone IMAGE:2002977 3',
            mRNA sequence.
ACCESSION   AI223644
NID         g3806357
VERSION     AI223644.1  GI:3806357
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT
            Contact: Robert Strausberg, Ph.D.
            Tel: (401) 496-1550
            Email: Robert_Strausberg@nih.gov
            Unknown library type
            Insert Length: 801  Std Error: 0.00
```

Seq primer: -40UP from Gibco.

```
FEATURES             Location/Qualifiers
     source          1. .216
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="17q25"
                     /clone="IMAGE:2002977"
                     /clone_lib="NCI_CGAP_Ov34"
                     /sex="female"
                     /tissue_type="borderline ovarian carcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP1; mRNA made from
                     borderline ovarian carcinoma, cDNA made by oligo-dT
                     priming  Directionally cloned  Size-selected on agarose
                     gel, average insert size 500 bp.  Primary library,
                     non-amplified.  "
BASE COUNT       73 a     58 c     53 g     32 t
ORIGIN
```

```
  Query Match             9.1%;  Score 173;  DB 43;  Length 216;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-77;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1722 acagaagggcagtgactcgacaaaggccacaggcagtccaggcctctctctgctccatcc 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10 ACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCC 69

Qy    1782 ccctgcctcccattctgcaccacacctggcatggtgcagggagacatctgcaccctgag 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 CCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCTGAG 129

Qy    1842 ttgggcagccaggagtgccccccgggaatggataataaagatactagagaaactg 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 TTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 182
```

```
RESULT   8
AI270872
LOCUS       AI270872      223 bp    mRNA            EST       17-NOV-1998
DEFINITION  qw67c10.x1 NCI_CGAP_Ov33 Homo sapiens cDNA clone IMAGE:1996146 3',
            mRNA sequence.
ACCESSION   AI270872
NID         g3890039
VERSION     AI270872.1  GI:3890039
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 223)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: I M A G E  Consortium, LLNL
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I M A G E  Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .223
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /clone "IMAGE:1996146"
                     /clone_lib="NCI_CGAP_Ov33"
                     /sex "female"
                     /tissue_type="borderline ovarian carcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP1; mRNA made from
                     borderline ovarian carcinoma, cDNA made by oligo-dT
                     priming. Directionally cloned. Size-selected on agarose
                     gel, average insert size 500 bp. Primary library,
                     non-amplified. "
BASE COUNT       78 a     54 c     54 g     32 t
ORIGIN


  Query Match             9.1%;  Score 173;  DB 44;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-77;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1722 acagaaggggcagtgactcgacaaaggccacaggcagtccaggcctctctctgctccatcc 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10 ACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCC 69

Qy    1782 cctgcctcccattctgcaccacacctggcatggtgcagggagacatctgcacccctgag 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 CCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAG 129

Qy    1842 ttgggcagccaggagtgccccccgggaatggataataaagatactagagaaactg 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 TTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 182


RESULT  9
AI271046
LOCUS       AI271046      222 bp    mRNA            EST       17-NOV-1998
DEFINITION  qw51q09.x1 NCI_CGAP_Br13 Homo sapiens cDNA clone IMAGE:1994656 3',
            mRNA sequence.
ACCESSION   AI271046
NID         g3899213
VERSION     AI271046.1  GI:3899213
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .222
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1994656"
                     /clone_lib="NCI_CGAP_Br13"
                     /sex="female"
                     /tissue_type="breast carcinoma in situ"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     carcinoma, cDNA made by oligo-dT priming. Directionally
                     cloned. Size-selected on agarose gel, average insert
                     size 500 bp. Primary library, non-amplified. "
BASE COUNT       76 a     58 c     55 g     33 t
ORIGIN


  Query Match             9.1%;  Score 173;  DB 44;  Length 222;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-77;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1722 acagaaggggcagtgactcgacaaaggccacaggcagtccaggcctctctctgctccatcc 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10 ACAGAAGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCC 69

Qy    1782 cctgcctcccattctgcaccacacctggcatggtgcagggagacatctgcacccctgag 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 CCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAG 129

Qy    1842 ttgggcagccaggagtgccccccgggaatggataataaagatactagagaaactg 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 TTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 182


RESULT  10
AI275319
LOCUS       AI275319      244 bp    mRNA            EST       21-DEC-1998
DEFINITION  qw66a10.x1 NCI_CGAP_Ov33 Homo sapiens cDNA clone IMAGE:1996026 3',
            mRNA sequence.
ACCESSION   AI275319
NID         g3897593
VERSION     AI275319.1  GI:3897593
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1899898.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 311   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1996026"
                     /clone_lib="NCI_CGAP_Ov33"
                     /sex="female"
                     /tissue_type="borderline ovarian carcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP1; mRNA made from
                     borderline ovarian carcinoma, cDNA made by oligo-dT
                     priming. Directionally cloned. Size-selected on agarose
                     gel, average insert size 500 bp. Primary library,
                     non-amplified. "
BASE COUNT       84 a     65 c     57 g     38 t
```

```
ORIGIN

  Query Match              9.1%;  Score 173;  DB 44;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-77;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1722 acagaagggcagtgactcgacaaagccacaggcagtccaggcctctctctgctccatcc 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACAGAAGGGCAGTGACTCGACAAAGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCC 60

Qy    1782 cctgcctcccattctgcaccacacctggcatggtgcagggagacatctgcacccctgag 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAG 120

Qy    1842 ttgggcagccaggagtgccccccgggaatggataataaagatactagagaaactg 1894
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 173


RESULT  11
AI275321
LOCUS       AI275321      232 bp    mRNA            EST       20-NOV-1998
DEFINITION  qw66a12.x1 NCI_CGAP_Ov33 Homo sapiens cDNA clone IMAGE:1996030 3',
            mRNA sequence.
ACCESSION   AI275321
NID         g3897595
VERSION     AI275321.1  GI:3897595
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 232)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1893942.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..232
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1996030"
                     /clone_lib="NCI_CGAP_Ov33"
                     /sex="female"
                     /tissue_type="borderline ovarian carcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: ovary; Vector: pAMP1; mRNA made from
                     borderline ovarian carcinoma, cDNA made by oligo-dT
                     priming.  Directionally cloned.  Size-selected on agarose
                     gel, average insert size 500 bp.  Primary library,
                     non-amplified.  "
BASE COUNT       74 a     61 c     57 g     35 t
ORIGIN

  Query Match              9.1%;  Score 173;  DB 44;  Length 232;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-77;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1722 acagaagggcagtgactcgacaaagccacaggcagtccaggcctctctctgctccatcc 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19 ACAGAAGGGCAGTGACTCGACAAAGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCC 78

Qy    1782 cctgcctcccattctgcaccacacctggcatggtgcagggagacatctgcacccctgag 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      79 CCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAG 138

Qy    1842 ttgggcagccaggagtgccccccgggaatggataataaagatactagagaaactg 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 TTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 191


RESULT  12
AI284816
LOCUS       AI284816      224 bp    mRNA            EST       21-DEC-1998
DEFINITION  qp25b04 x1 NCI_CGAP_Br12 Homo sapiens cDNA clone IMAGE:1965775,
            mRNA sequence.
ACCESSION   AI284816
NID         g3923049
VERSION     AI284816.1  GI:3923049
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 224)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797188.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 296   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..224
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="12p"
                     /clone="IMAGE:1965775"
                     /clone_lib="NCI_CGAP_Br12"
                     /sex="female"
                     /tissue_type="invasive carcinoma"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     carcinoma, cDNA made by oligo-dT priming.
                     Non-directionally cloned. Size-selected on agarose gel,
                     average insert  size 500 bp.  Non-amplified library.  "
BASE COUNT       80 a     59 c     53 g     32 t
ORIGIN

  Query Match              9.1%;  Score 173;  DB 44;  Length 224;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-77;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1722 acagaagggcagtgactcgacaaagccacaggcagtccaggcctctctctgctccatcc 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10 ACAGAAGGGCAGTGACTCGACAAAGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCC 69
```

```
Qy    1787 ccctgcctcccattctgcaccacacctggcatggtgcagggagacatctgcacccctgag 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 CCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAG 129

Qy    1842 ttgggcagccaggagtgccccccgggaatggataataaagatactagagaaactg 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 TTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 182


RESULT  14
AI289687
LOCUS       AI289687      199 bp    mRNA            EST       30-NOV-1998
DEFINITION  qw42e08.x1 NCI_CGAP_Br13 Homo sapiens cDNA clone IMAGE:1993766 3',
            mRNA sequence.
ACCESSION   AI289687
NID         g3931528
VERSION     AI289687.1  GI:3931528
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..199
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1993766"
                     /clone_lib="NCI_CGAP_Br13"
                     /sex="female"
                     /tissue_type="breast carcinoma in situ"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     carcinoma; cDNA made by oligo-dT priming. Directionally
                     cloned. Size-selected on agarose gel, average insert
                     size 500 bp. Primary library, non-amplified. "
BASE COUNT       62 a     57 c     50 g     30 t
ORIGIN


  Query Match          9.1%;  Score 173;  DB 44;  Length 199;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-77;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1722 acagaaggggcagtgactcgacaaaggccacaggcagtccaggcctctctctgctccatcc 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10 ACAGAAGGGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCC 69

Qy    1782 ccctgcctcccattctgcaccacacctggcatggtgcagggagacatctgcacccctgag 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 CCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAG 129

Qy    1842 ttgggcagccaggagtgccccccgggaatggataataaagatactagagaaactg 1894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 TTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 182


RESULT  15
AI305465
LOCUS       AI305465      218 bp    mRNA            EST       03-FEB-1999
```

```
DEFINITION  qw46e09.x1 NCI_CGAP_Br13 Homo sapiens cDNA clone IMAGE:1994150 3',
            mRNA sequence.
ACCESSION   AI305465
NID         g3990356
VERSION     AI305465.1  GI:3990356
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata, Craniata, Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 218)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html

            Insert Length: 298   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1..218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1994150"
                     /clone_lib="NCI_CGAP_Br13"
                     /sex="female"
                     /tissue_type="breast carcinoma in situ"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: breast; Vector: pAMP1; mRNA made from breast
                     carcinoma; cDNA made by oligo-dT priming. Directionally
                     cloned. Size-selected on agarose gel, average insert
                     size 500 bp. Primary library, non-amplified. "
BASE COUNT       75 a     58 c     53 g     32 t
ORIGIN


  Query Match             9.1%;  Score 173;  DB 44;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-77;
  Matches  173;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1722 acaaaaggcagtgactcgacaaaggccacaggcagtccaggcctctctctgctccatcc 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10 ACAGAAGGCAGTGACTCGACAAAGGCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCC 69

Qy    1782 cccgccctcccattctgcaccacacctggcatggtgcagggagacatctgcacccctgag 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 CCCTGCCTCCCATTCTGCACCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAG 129

Qy    1842 ttgggcagccaggagtgcccccgggaatggataataaagatactagagaactg 1894
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 TTGGGCAGCCAGGAGTGCCCCCGGGAATGGATAATAAAGATACTAGAGAACTG 182


Search completed: November 22, 1999, 18:05:22
Job time: 445 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd

OM protein - protein search, using sw model

Run on: November 20, 1999, 13:25:12 ; Search time 15.75 Seconds
(without alignments)
1309.023 Million cell updates/sec

Title: US-09-126-945-2
Perfect score: 335
Sequence: 1 MGSASPGLSSVSPSHLLLPP . . . GIIRKPDISQRLVYQFVHPI 335

Scoring table: OLIGO

Searched: 201082 seqs, 61543640 residues

Database: SPTREMBL_10:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	335	100.0	335	4	O95238	O95238 homo sapien
2	21	6.3	363	5	O97145	O97145 strongyloce
3	14	4.2	454	4	O75849	O75849 homo sapien
4	14	4.2	494	13	O57586	O57586 brachydanio
5	14	4.2	477	13	O93322	O93322 xenopus lae
6	14	4.2	236	13	Q9YHW7	Q9yhw7 gallus gall
7	14	4.2	477	13	Q9YHW6	Q9yhw6 gallus gall
8	14	4.2	336	13	Q9YHW5	Q9yhw5 gallus gall
9	9	2.7	532	5	Q19695	Q19695 caenorhabdi
10	8	2.4	472	1	P95453	P95453 pyrococcus
11	8	2.4	385	4	Q14319	Q14319 homo sapien
12	8	2.4	254	4	Q16199	Q16199 homo sapien
13	8	2.4	196	4	Q16203	Q16203 homo sapien
14	8	2.4	258	4	Q99581	Q99581 homo sapien
15	8	2.4	95	4	O95058	O95058 homo sapien
16	8	2.4	33	5	Q25573	Q25573 unidentifie
17	8	2.4	33	5	Q26672	Q26672 tethya aura
18	8	2.4	33	5	Q26673	Q26673 tethya aura
19	8	2.4	208	5	Q18579	Q18579 caenorhabdi
20	8	2.4	441	5	Q27378	Q27378 caenorhabdi
21	8	2.4	2219	5	Q23388	Q23388 caenorhabdi
22	8	2.4	343	11	Q70132	Q70132 rattus norv
23	8	2.4	471	13	O57566	O57566 carassius a
24	8	2.4	492	13	Q93425	Q93425 petromyzon
25	7	2.1	135	2	P95327	P95327 myxococcus
26	7	2.1	141	2	Q53755	Q53755 streptomyce
27	7	2.1	361	2	P94521	P94521 bacillus su
28	7	2.1	754	2	Q34740	Q34740 bacillus su
29	7	2.1	147	2	P96272	P96272 mycobacteri
30	7	2.1	238	2	O53500	O53500 mycobacteri
31	7	2.1	452	2	O31855	O31855 bacillus su
32	7	2.1	512	2	O67911	O67911 aquifex aeo
33	7	2.1	564	2	O06059	O06059 mycobacteri
34	7	2.1	575	2	O07433	O07433 mycobacteri
35	7	2.1	600	2	O86572	O86572 streptomyce
36	7	2.1	517	2	O86651	O86651 streptomyce
37	7	2.1	686	2	O87092	O87092 bacillus su
38	7	2.1	722	2	O84734	O84734 chlamydia p
39	7	2.1	1338	3	O60030	O60030 kluyveromyc
40	7	2.1	820	4	O60585	O60585 homo sapien
41	7	2.1	578	4	Q92839	Q92839 homo sapien
42	7	2.1	133	4	Q08124	Q08124 homo sapien
43	7	2.1	581	4	Q15723	Q15723 homo sapien
44	7	2.1	533	4	Q15724	Q15724 homo sapien
45	7	2.1	490	13	Q9YH24	Q9yh24 brachydanio

ALIGNMENTS

RESULT 1
O95238
ID O95238 PRELIMINARY; PRT; 335 AA.
AC O95238;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ETS TRANSCRIPTION FACTOR PDEF.
GN PDEF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA OETTGEN P., FINGER E., AKBARALI Y., THAMRONGSAK U., DUBE A.,
RA BOLTAX J., BROWN L., KUNSCH C., LIBERMANN T.A.;
RT "Isolation and characterization of a novel prostate
RT epithelium-specific Ets transcription factor, PDEF.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF071538; AAC95296.1; -.
SQ SEQUENCE 335 AA; 37517 MW; 874A5DE8 CRC32;

Query Match 100.0%; Score 335; DB 4; Length 335;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWSPSPPATPEQGLSAFYL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWSPSPPATPEQGLSAFYL 60

Qy      61 SYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPAGSLDLVPGGLTLEEHSLE 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPAGSLDLVPGGLTLEEHSLE 120

Qy     121 QVQSMVVGEVLKDIETACKLLNIIADPMDWSPSNVQKWLLWIEHQYRLPPMGKAFQELAG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QVQSMVVGEVLKDIETACKLLNIIADPMDWSPSNVQKWLLWIEHQYRLPPMGKAFQELAG 180

Qy     181 KELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKERTSPGAIHYCASTSEESWTDSEV 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 KELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKERTSPGAIHYCASTSEESWTDSEV 240

Qy     241 DSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEKGIFKIEDSAQVAPLWGIPRNRPAM 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 DSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEKGIFKIEDSAQVAPLWGIPRNRPAM 300

Qy     301 NYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHPI 335
           |||||||||||||||||||||||||||||||||||
Db     301 NYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHPI 335
```

RESULT 2
Q97134
ID Q97134 PRELIMINARY; PRT; 363 AA.
AC Q97134;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ETS34 TRANSCRIPTION FACTOR.
GN ETS34.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RA WEI Z., ANGERER R.C., ANGERER L.M.;
RT "Identification of a new sea urchin ets protein, SpEts4, by yeast
RT one-hybrid screening with the hatching enzyme promoter.";
RL Mol. Cell. Biol. 20:63(1999).
DR EMBL; AF100964; AAD19942.1; -.
SQ SEQUENCE 363 AA; 40768 MW; C4B9BDFA CRC32;

Query Match 4.2%; Score 14; DB 5; Length 363;
Best Local Similarity 100.0%; Pred. No. 3.3e-12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 294 KNRPAMNYDKLSRS 307
||||||||||||||
Db 62 KNRPAMNYDKLSRS 75

RESULT 3
O75849
ID O75849 PRELIMINARY; PRT; 454 AA.
AC O75849;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ETS RELATED TRANSCRIPTION FACTOR (BINDS CGGAW).
GN ETV1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA ZERSKY R., EDWARDS J.;
RT "The sequence of Homo sapiens PAC clone DJ0685A02.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.H.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC004857; AAC02435.1; -.
DR PFAM; PF00178; Ets; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
SQ SEQUENCE 454 AA; 52474 MW; 63D32BF9 CRC32;

Query Match 4.2%; Score 14; DB 4; Length 454;
Best Local Similarity 100.0%; Pred. No. 5.4e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 379 KNRPAMNYDKLSRS 392

RESULT 4
O57586
ID O57586 PRELIMINARY; PRT; 494 AA.
AC O57586;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ETS-DOMAIN TRANSCRIPTION FACTOR PEA3.
GN PEA3.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinoidea; Cyprinidae; Rasborinae; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RA BROWN L.A., AMORES A., SCHILLING T.F., JOWETT T., BAERT J.L.,
RA DE LAUNOIT Y., SHARROCKS A.D.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ003200; CAA05980.1; -.
DR PFAM; PF00178; Ets; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR ZFIN; ZDB-GENE-990415-71; pea3.
SQ SEQUENCE 494 AA; 55620 MW; 249CBBDC CRC32;

Query Match 4.2%; Score 14; DB 13; Length 494;
Best Local Similarity 100.0%; Pred. No. 5.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 394 KNRPAMNYDKLSRS 407

RESULT 5
O93322
ID O93322 PRELIMINARY; PRT; 477 AA.
AC O93322;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ETS-RELATED TRANSCRIPTION FACTOR XER81 (FRAGMENT).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RA CHEN J.L., GRUNZ H., PANITZ F., PIELER T., HOLLEMANN T.;
RT "Xer81 - A ets-related gene has a function during early Xenopus
RT development.";
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF057565; AAC25157.1; -.
DR PFAM; PF00178; Ets; 1.
FT NON_TER 477 477
SQ SEQUENCE 477 AA; 55124 MW; B894A575 CRC32;

Query Match 4.2%; Score 14; DB 13; Length 477;
Best Local Similarity 100.0%; Pred. No. 5.7e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 379 KNRPAMNYDKLSRS 392

RESULT 6
Q9YHW7
ID Q9YHW7 PRELIMINARY; PRT; 236 AA.
AC Q9YHW7;

DT 1-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 1-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ETS DOMAIN PROTEIN (FRAGMENT).
GN ERM.
OS Gallus gallus (Chicken)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus
RN [1]
RP SEQUENCE FROM N.A.
RA LIN J.H., SAITO T., ANDERSON D.J., LANCE-JONES C., JESSELL T.M.,
RA ARBER S.;
RT "Functionally related motor neuron pool and muscle sensory afferent
RT subtypes defined by coordinate ETS gene expression.";
RL Cell 95:393-407(1998).
DR EMBL; AF075706; AAC97201.1; -.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
FT NON_TER 1 1
SQ SEQUENCE 236 AA; 27427 MW; 5E10959D CRC32;

Query Match 4.2%; Score 14; DB 13; Length 236;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 138 KNRPAMNYDKLSRS 151

RESULT 7
Q9YHW6
ID Q9YHW6 PRELIMINARY; PRT; 477 AA.
AC Q9YHW6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 1-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ETS DOMAIN PROTEIN.
GN ERM.
OS Gallus gallus (Chicken)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus
RN [1]
RP SEQUENCE FROM N.A.
RA LIN J.H., SAITO T., ANDERSON D.J., LANCE-JONES C., JESSELL T.M.,
RA ARBER S.;
RT "Functionally related motor neuron pool and muscle sensory afferent
RT subtypes defined by coordinate ETS gene expression.";
RL Cell 95:393-407(1998).
DR EMBL; AF075707; AAC97202.1; -.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
SQ SEQUENCE 477 AA; 55199 MW; D6887E34 CRC32;

Query Match 4.2%; Score 14; DB 13; Length 477;
Best Local Similarity 100.0%; Pred. No. 5.7e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 479 KNRPAMNYDKLSRS 492

RESULT 8
Q9YHW5
ID Q9YHW5 PRELIMINARY; PRT; 336 AA.
AC Q9YHW5;
DT 1-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ETS DOMAIN PROTEIN (FRAGMENT).
GN PEA3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RA LIN J.H., SAITO T., ANDERSON D.J., LANCE-JONES C., JESSELL T.M.,
RA ARBER S.;
RT "Functionally related motor neuron pool and muscle sensory afferent
RT subtypes defined by coordinate ETS gene expression.";
RL Cell 95:393-407(1998).
DR EMBL; AF075708; AAC97203.1; -.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
FT NON_TER 1 1
SQ SEQUENCE 336 AA; 38518 MW; 9C822D20 CRC32;

Query Match 4.2%; Score 14; DB 13; Length 336;
Best Local Similarity 100.0%; Pred. No. 4.2e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 238 KNRPAMNYDKLSRS 251

RESULT 9
Q19695
ID Q19695 PRELIMINARY; PRT; 532 AA.
AC Q19695;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE SIMILAR TO DNA BINDING PROTEIN.
GN F22A3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA HALLSWORTH K.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U41547; AAA83194.1; -.
DR PFAM; PF00178; Ets; 1.
SQ SEQUENCE 532 AA; 60198 MW; 9806767E CRC32;

Query Match 2.7%; Score 9; DB 5; Length 532;
Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSRS 308

```
                 ||||||||
Db     145 MNYDKLSRS 513


RESULT  9
O59453
ID   O59453      PRELIMINARY;      PRT:   472 AA.
AC   O59453;
DT   1-APR-1998 (TrEMBLrel. 07, Created)
DT   1-APR-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   412AA LONG HYPOTHETICAL PROTEIN.
GN   PH1780.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA   SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA   FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., ROBB F.T., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.;
RT   "Complete sequence and gene organization of the genome of a
RT   hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000007; BAA29008.1; -.
SQ   SEQUENCE   472 AA;  51619 MW;  FC189374 CRC32;


  Query Match              2.4%;  Score 8;  DB 1;  Length 472;
  Best Local Similarity   100.0%;  Pred. No. 7.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     273 NKEKGIFK 280
           ||||||||
Db      42 NKEKGIFK 49


RESULT  11
Q14439
ID   Q14439      PRELIMINARY;      PRT:   385 AA.
AC   Q14439;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   FLI-1.
GN   FLI-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE; 94176799.
RA   THOMAS R., MAY W., DENNY C., PASKIND M., MOORE J., MAKI R.A., BECK E.,
RA   KIEMSY M.;
RT   "Human FLI-1 localizes to chromosome 11Q24 and has an aberrant
RT   transcript in neuroepithelioma.";
RL   Biochim. Biophys. Acta 1172:155-158(1993).
DR   EMBL; M94255; AAA58480.1; -.
DR   PFAM; PF00178; Ets; 1.
DR   PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR   PROSITE; PS00346; ETS_DOMAIN_2; 1.
SQ   SEQUENCE   385 AA;  43437 MW;  0820C243 CRC32;


  Query Match              2.4%;  Score 8;  DB 4;  Length 385;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     300 MNYDKLSR 307
           ||||||||
Db     264 MNYDKLSR 271


RESULT  12
Q16199
ID   Q16199      PRELIMINARY;      PRT:   254 AA.
AC   Q16199;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   EWS-ERG FUSION PROTEIN TYPE 1E (FRAGMENT).
GN   EWS-ERG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94314948
RA   GIOVANNINI M., BIEGEL J.A., SERRA M., WANG J.Y., WEI Y.H., NYCUM L.,
RA   EMANUEL B.S., EVANS G.A.;
RT   "EWS-erg and EWS-Fli1 fusion transcripts in Ewing's sarcoma and
RT   primitive neuroectodermal tumors with variant translocations.";
RL   J. Clin. Invest. 94:489-496(1994).
DR   EMBL; S72621; AAB31417.1; -.
DR   PFAM; PF00178; Ets; 1.
DR   PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR   PROSITE; PS00346; ETS_DOMAIN_2; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   254 AA;  28577 MW;  596580CD CRC32;


  Query Match              2.4%;  Score 8;  DB 4;  Length 254;
  Best Local Similarity   100.0%;  Pred. No. 4.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     300 MNYDKLSR 307
           ||||||||
Db     140 MNYDKLSR 147


RESULT  13
Q16203
ID   Q16203      PRELIMINARY;      PRT:   196 AA.
AC   Q16203;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   EWS     EWS-ERG=EWS-ERG FUSION PROTEIN TYPE 9E (FRAGMENT).
GN   EWS-ERG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94314948.
RA   GIOVANNINI M., BIEGEL J.A., SERRA M., WANG J.Y., WEI Y.H., NYCUM L.,
RA   EMANUEL B.S., EVANS G.A.;
RT   "EWS-erg and EWS-Fli1 fusion transcripts in Ewing's sarcoma and
RT   primitive neuroectodermal tumors with variant translocations.";
RL   J. Clin. Invest. 94:489-496(1994).
DR   EMBL; S72865; AAB31419.1; -.
DR   PFAM; PF00178; Ets; 1.
DR   PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR   PROSITE; PS00346; ETS_DOMAIN_2; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   196 AA;  22268 MW;  FBCD632E CRC32;


  Query Match              2.4%;  Score 8;  DB 4;  Length 196;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      [illegible] MNYDKLSR 307
                    ||||||||
Db      [illegible] MNYDKLSR 89


RESULT  14
Q995[illegible]
ID   [illegible]      PRELIMINARY;      PRT;   238 AA.
AC   [illegible];
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   FEV PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97236656.
RA   PETER, C.; COUTURIER J., PACQUEMENT H., MICHON J., THOMAS G.,
RA   MAGHELENA; H., DELAITRE O.;
RT   "A new member of the ETS family fused to EWS in Ewing tumors.";
RL   Oncogene 14:1159-1164(1997).
DR   EMBL; Y08976; CAA70169.1; -.
DR   PFAM; PF00178; Ets; 1.
DR   PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR   PROSITE; PS00346; ETS_DOMAIN_2; 1.
SQ   SEQUENCE   238 AA;  25030 MW;  D4057DD0 CRC32;


  Query Match             2.4%;  Score 8;  DB 4;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      [illegible] MNYDKLSR 307
                    ||||||||
Db      [illegible] MNYDKLSR 103


RESULT  15
Q95[illegible]
ID   [illegible]      PRELIMINARY;      PRT;    95 AA.
AC   [illegible];
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ELK VARIANT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ARYEE D.N.T., KOVAR H.;
RT   "Novel family members HuER71, ELFR, and ELKv among ETS-related genes
RT   coexpressed with EWS-FLI1 in Ewing tumor cell lines.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF[illegible]672; AAD00862.1; -.
DR   PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR   PROSITE; PS00346; ETS_DOMAIN_2; 1.
SQ   SEQUENCE   95 AA;  11217 MW;  B661ECA2 CRC32;


  Query Match             2.4%;  Score 8;  DB 4;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      [illegible] MNYDKLSR 307
                    ||||||||
Db      [illegible] MNYDKLSR 62
```

Search completed: November 20, 1999, 13:29:16
Job time: 244 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 18:07:15 ; Search time 11.23 Seconds
(without alignments)
1195.186 Million cell updates/sec

Title: US-09-126-945-2
Perfect score: 335
Sequence: 1 MGSASPGLSSVSPSHLLLPP . . GIIPKPDISQPLVYQFVHPI 335

Scoring table: OLIGO

Searched: 122810 seqs, 40065486 residues

Database: PIR_6u:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24	7.2	114	2	S28821	transcription fact
2	14	4.2	477	1	I38893	transcription fact
3	14	4.2	477	1	B46396	transcription fact
4	14	4.2	510	1	S43692	transcription fact
5	14	4.2	462	1	S35534	adenovirus E1A enh
6	14	4.2	555	1	S24061	transcription fact
7	10	3.0	453	2	B42093	serum response fac
8	10	3.0	430	2	I48755	mSAP1a - mouse
9	10	3.0	405	2	A42093	serum response fac
10	8	2.4	486	1	TVHUEG	transforming prote
11	8	2.4	478	1	S60754	transcription fact
12	8	2.4	452	1	I37565	transforming prote
13	8	2.4	385	1	S29844	transforming prote
14	8	2.4	452	1	S17403	transforming prote
15	8	2.4	453	1	S49019	transforming prote
16	8	2.4	428	1	TVHUEK	transforming prote
17	8	2.4	102	2	S28923	transcription fact
18	8	2.4	94	2	S28920	transcription fact
19	8	2.4	172	2	A56646	transcription fact
20	8	2.4	272	2	A54617	transcription fact
21	8	2.4	179	2	S51226	transcription fact
22	8	2.4	420	2	JC4965	elk1 protein - mou
23	8	2.4	409	2	I48680	transcription fact
24	8	2.4	407	2	I38062	Net - human
25	8	2.4	247	2	S72775	probable glucose-6
26	8	2.4	924	2	A24692	ecdysone-induced p
27	8	2.4	883	2	B34692	ecdysone-induced p
28	8	2.4	807	2	A53225	ecdysone-induced p
29	8	2.4	973	2	B53225	ecdysone-induced p
30	8	2.4	548	2	S59133	ETS2 repressor fac
31	8	2.4	250	2	A54308	PE-1 protein - hum
32	8	2.4	472	2	E71189	hypothetical prote
33	7	2.1	421	1	DEHUCM	acyl-CoA dehydroge
34	7	2.1	421	1	DERTCM	acyl-CoA dehydroge
35	7	2.1	444	1	B41864	acetyl-CoA--deacet
36	7	2.1	1748	1	JQ1555	genome polyprotein
37	7	2.1	2077	1	WZBEZ4	240K tegument prot
38	7	2.1	421	2	A55724	acyl-CoA dehydroge
39	7	2.1	270	2	S47099	site-specific DNA-
40	7	2.1	361	2	A69984	endo-1,4-beta-gluc
41	7	2.1	575	2	H70906	probable ilvD prot
42	7	2.1	686	2	A44842	cGMP-gated ion cha
43	7	2.1	133	2	I37382	cGMP-gated cation
44	7	2.1	521	2	S54266	glycoprotein gC -
45	7	2.1	411	2	A49127	homeotic protein A

ALIGNMENTS

RESULT 1
S28821
transcription factor ets-98B - fruit fly (Drosophila melanogaster) (fragment)
N;Alternate names: transforming protein ets-4
C;Species: Drosophila melanogaster
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 24-Sep-1998
C;Accession: S28821
R;Chen, T.; Bunting, M.; Karim, F.D.; Thummel, C.S.
Dev. Biol. 151, 176-191, 1992
A;Title: Isolation and characterization of five Drosophila genes that encode an ets-r
A;Reference number: S28819; MUID:92249640
A;Accession: S28821
A;Molecule type: DNA
A;Residues: 1-114 <CHE>
A;Cross-references: EMBL:M88474; NID:g157195; PID:g157196
C;Genetics:
A;Gene: Ets98B
A;Cross-references: FlyBase:FBgn0005659
C;Superfamily: ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;28-109/Domain: ets DNA-binding domain homology <ETS>

Query Match 7.2%; Score 24; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 3.2e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 294 PKNPPAMNYDKLSPSIRQYYKKGI 317
||||||||||||||||||||||||
Db 71 PKNPPAMNYDKLSPSIRQYYKKGI 94

RESULT 2
I38893
transcription factor ER81 - human
N;Alternate names: ets translocation variant 1
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 23-May-1997 #text_change 18-Sep-1998
C;Accession: I38893; S61541
R;Jeon, I.S.; Davis, J.N.; Braun, B.S.; Sublett, J.E.; Roussel, M.F.; Denny, C.T.; Sh
Oncogene 10, 1229-1234, 1995
A;Title: A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS
A;Reference number: I38893; MUID:95215084
A;Accession: I38893
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-477 <RES>
A;Cross-references: EMBL:U17163; NID:g596005; PID:g596006
R;Monte, D.; Coutte, L.; Baert, J.L.; Angeli, I.; Stehelin, D.; de Launoit, Y.
Oncogene 11, 771-779, 1995
A;Title: Molecular characterization of the ets-related human transcription factor ER8
A;Reference number: S61541; MUID:95380185
A;Accession: S61541
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-38,'V',40-60,'V',80-116,'G',118-126,'K',128-252,254-348,'A',350-477 <MO
A;Cross-references: EMBL:X87175; NID:g1045060; PID:g1045061
C;Genetics:
A;Gene: GDB:ETV1
A;Cross-references: GDB:335229; OMIM:600541
A;Map position: 21q22.3-21q22.3
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology

C;Keywords: DNA binding; nucleus; transcription factor
F;347-415/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 477;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 479 KNRPAMNYDKLSRS 492

RESULT 4
B46306
transcription factor ER81 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 15-May-1994 #sequence_revision 23-May-1997 #text_change 05-Sep-1997
C;Accession: B46306
R;Brown, T.A.; McKnight, S.L.
Genes Dev. 6, 2502-2512, 1992
A;Title: Specificities of protein-protein and protein-DNA interaction of GABPalpha and t
A;Reference number: A46306
A;Accession: B46306
A;Molecule type: mRNA
A;Residues: 1-477 <BRO>
A;Cross-references: GB:L10426; NID:g193195; PID:g515964
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;347-415/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 477;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 479 KNRPAMNYDKLSRS 492

RESULT 5
S43692
transcription factor ERM - human
N;Alternate names: ets variant gene 5; ets-related molecule
C;Species: Homo sapiens (man)
C;Date: 24-Jan-1995 #sequence_revision 23-May-1997 #text_change 18-Sep-1998
C;Accession: S43692
R;Monte, D.; Baert, J.L.; Defossez, P.A.; de Launoit, Y.; Stehelin, D.
Oncogene 9, 1397-1406, 1994
A;Title: Molecular cloning and characterization of human ERM, a new member of the Ets fa
A;Reference number: S43692; MUID:94203669
A;Accession: S43692
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-510 <MON>
A;Cross-references: EMBL:X76184; NID:g479166; PID:g479167
C;Genetics:
A;Gene: GDB:ETV5; ERM; ETV2LL
A;Cross-references: GDB:361087; OMIM:601600
A;Map position: 3q27-3q29
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;347-448/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 512 KNRPAMNYDKLSRS 525

RESULT 5
S35534
adenovirus E1A enhancer-binding protein E1AF - human (fragment)
N;Alternate names: ets translocation variant 4
C;Species: Homo sapiens (man)
C;Date: 09-Dec-1993 #sequence_revision 23-May-1997 #text_change 18-Sep-1998
C;Accession: S35534
R;Higashino, F.; Yoshida, K.; Fujinaga, Y.; Kamio, K.; Fujinaga, K.
Nucleic Acids Res. 21, 547-553, 1993
A;Title: Isolation of a cDNA encoding the adenovirus E1A enhancer binding protein: a
A;Reference number: S35534; MUID:93181246
A;Accession: S35534
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-462 <HIG>
A;Cross-references: GB:D12765; NID:g216516; PID:g1002724; PID:g219611
C;Genetics:
A;Gene: GDB:ETV4; E1A-F
A;Cross-references: GDB:512230; OMIM:600711
A;Map position: 17q21-17q21
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;321-399/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 363 KNRPAMNYDKLSRS 376

RESULT 6
S24061
transcription factor PEA3 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Feb-1994 #sequence_revision 23-May-1997 #text_change 05-Sep-1997
C;Accession: S24061
R;Xin, J.H.; Cowie, A.; Lachance, P.; Hassell, J.A.
Genes Dev. 6, 481-496, 1992
A;Title: Molecular cloning and characterization of PEA3, a new member of the Ets onco
A;Reference number: S24061; MUID:92192459
A;Accession: S24061
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-555 <XIN>
A;Cross-references: EMBL:X63190; NID:g53627; PID:g53628
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;414-492/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 555;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 456 KNRPAMNYDKLSRS 469

RESULT 7
B42093
serum response factor accessory protein-1 form a (SAP-1a) - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Aug-1998
C;Accession: B42093; S28817
R;Dalton, S.; Treisman, R.
Cell 68, 597-612, 1992

A;Title: Characterization of SAP-1, a protein recruited by serum response factor to the
A;Reference number: A42093; MUID:92154673
A;Accession: A42093
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-453 <DAL>
A;Cross-references: EMBL:M85165
A;Experimental source: HeLa cells
A;Note: sequence extracted from NCBI backbone (NCBIN:82523, NCBIP:82524)
C;Superfamily: elk-1 transforming protein; ets DNA-binding domain homology
C;Keywords: DNA binding
F;7-85/Domain: ets DNA-binding domain homology <ETS>

Query Match 3.0%; Score 10; DB 2; Length 453;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 VARLWGIRKN 296
||||||||||
Db 41 VARLWGIRKN 50

RESULT 9
I48755
mSAP1a - mouse
C;Species: Mus musculus (house mouse)
C;Date: 2-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Aug-1998
C;Accession: I48755
R;Giampuzzi, A.; Pintzas, A.; Maira, S.M.; Sobieszczuk, P.; Wasylyk, B.
Genes Dev. 8, 1502-1513, 1994
A;Title: Net, a new ets transcription factor that is activated by Ras
A;Reference number: A53837; MUID:95047310
A;Accession: I48755
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-430 <RES>
A;Cross-references: EMBL:Z34885; NID:g535127; PID:g535128
C;Superfamily: elk-1 transforming protein; ets DNA-binding domain homology
F;7-85/Domain: ets DNA-binding domain homology <ETS>

Query Match 3.0%; Score 10; DB 2; Length 430;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 VARLWGIRKN 296
||||||||||
Db 41 VARLWGIRKN 50

RESULT 10
A42093
serum response factor accessory protein-1 form b (SAP-1b) - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Aug-1998
C;Accession: A42093; S28816
R;Dalton, S.; Treisman, R.
Cell 68, 597-612, 1992
A;Title: Characterization of SAP-1, a protein recruited by serum response factor to the
A;Reference number: A42093; MUID:92154673
A;Accession: A42093
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-405 <DAL>
A;Cross-references: GB:M85164; NID:g338034; PID:g338035
A;Experimental source: HeLa cells
A;Note: sequence extracted from NCBI backbone (NCBIN:82525, NCBIP:82527)
C;Superfamily: elk-1 transforming protein; ets DNA-binding domain homology
C;Keywords: DNA binding
F;7-85/Domain: ets DNA-binding domain homology <ETS>

Query Match 3.0%; Score 10; DB 2; Length 405;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 VARLWGIRKN 296
||||||||||
Db 41 VARLWGIRKN 50

RESULT 10
TVHUEG
transforming protein erg-3 - human
N;Contains: transforming protein erg-1; transforming protein erg-2
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1989 #sequence_revision 30-May-1997 #text_change 31-Oct-1997
C;Accession: A94294; A94178; I58410; A28041; A29515
R;Rao, V.N.; Papas, T.S.; Shyam, E.; Reddy, P.
Science 237, 635-639, 1987
A;Title: erg, a human ets-related gene on chromosome 21: alternative splicing, polyad
A;Reference number: A94294; MUID:87263429
A;Accession: A94294
A;Molecule type: mRNA
A;Residues: 1-231,256-486 <REA>
A;Cross-references: GB:M17254; NID:g182186; PID:g182187
R;Reddy, E.S.P.; Rao, V.N.; Papas, T.S.
Proc. Natl. Acad. Sci. U.S.A. 84, 6131-6135, 1987
A;Title: The erg gene: a human gene related to the ets oncogene.
A;Reference number: A94178; MUID:87317608
A;Accession: A94178
A;Molecule type: mRNA
A;Residues: 100-231,256-486 <REB>
A;Cross-references: GB:M21535; NID:g182182; PID:g182185
R;Prasad, D.D.; Rao, V.N.; Lee, L.; Reddy, E.S.
Oncogene 9, 669-673, 1994
A;Title: Differentially spliced erg-3 product functions as a transcriptional activato
A;Reference number: I58410; MUID:94119611
A;Accession: I58410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 230-259 <RES>
A;Cross-references: GB:S68130; NID:g544960; PID:g544961
C;Genetics:
A;Gene: GDB:ERG
A;Cross-references: GDB:119884; OMIM:165080
A;Map position: 21q22.2-21q22.2
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C;Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription
F;1-231,256-486/Product: transforming protein erg-2 #status predicted <EG2>
F;100-231,256-486/Product: transforming protein erg-1 #status predicted <EG1>
F;126-209/Domain: ets RII regulatory region homology <ETS2>
F;320-398/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 367 MNYDKLSR 374

RESULT 11
S60754
transcription factor erg - chicken
C;Species: Gallus gallus (chicken)
C;Date: 27-Apr-1996 #sequence_revision 23-May-1997 #text_change 10-Jul-1998
C;Accession: S60754
R;Dhordain, P.; Dewitte, F.; Desbiens, X.; Stehelin, D.; Duterque-Coquillaud, M.
Mech. Dev. 50, 17-28, 1995
A;Title: Mesodermal expression of the chicken erg gene associated with precartilagino
A;Reference number: S60754; MUID:95329425
A;Accession: S60754

A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-478 <DBE>
A;Cross-references: EMBL:X77154; NID:g790439; PID:g790440
C;Genetics:
A;Gene: GDB:ERG
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regul
C;Keywords: DNA binding; phosphoprotein; proto-oncogene; transcription factor; transform
F;119-192/Domain: ets RII regulatory region homology <ETS2>
F;312-390/Domain: ets DNA binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 370 MNYDKLSR 377

RESULT 12
I47567
transforming protein fli, long splice form - human
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1996 #sequence_revision 30-May-1997 #text_change 20-Mar-1998
C;Accession: I47565; S29844; S35506; A49000; A49015; I54170
R;Delattre, O.; Zucman, J.; Plougastel, B.; Desmaze, C.; Melot, T.; Peter, M.; Kovar, H.
Nature 359, 162-165, 1992
A;Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocation i
A;Reference number: I47478; MUID:92396239
A;Accession: I47565
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-452 <RES>
A;Cross-references: EMBL:X67001; NID:g32529; PID:g32530
R;Bromas, R.; May, W.; Denny, C.; Paskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Klemsz,
Biochim. Biophys. Acta 1172, 155-158, 1993
A;Title: Human FLI-1 localizes to chromosome 11Q24 and has an aberrant transcript in neu
A;Reference number: S29844; MUID:93176799
A;Accession: S29844
A;Molecule type: mRNA
A;Residues: 1-68,'V',70-129,'A',131-132,'V',134-322,'Q',324-425,427-452 <HRO1>
A;Cross-references: EMBL:M93255
A;Note: the authors translated the codon GTG for residue 69 as Glu
R;Bromas, R.A.; May, W.; Denny, C.; Paskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Klemsz
submitted to the EMBL Data Library, May 1993
A;Description: Human FLI-1, an ETS oncogene family member preferentially expressed in he
A;Reference number: S35506
A;Accession: S35506
A;Molecule type: mRNA
A;Residues: 1-68,'V',70-76,78-129,'A',131-132,'V',134-322,'Q',324-425,427-452 <HPO2>
A;Cross-references: EMBL:M93255; NID:g182659; PID:g182660
R;Prasad, D.D.; Rao, V.N.; Reddy, E.S.
Cancer Res. 52, 5833-5837, 1992
A;Title: Structure and expression of human Fli-1 gene.
A;Reference number: A49000; MUID:93007976
A;Accession: A49000
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-322,'Q',324-452 <PRA>
A;Cross-references: GB:S45205; NID:g257353; PID:g257354
A;Note: sequence extracted from NCBI backbone (NCBIN:115336, NCBIP:115337)
R;Watson, D.K.; Smyth, F.E.; Thompson, D.M.; Cheng, J.Q.; Testa, J.R.; Papas, T.S.; Seth
Cell Growth Differ. 3, 705-714, 1992
A;Title: The ERGB/FLI-1 gene: isolation and characterization of a new member of the fami
A;Reference number: A49015; MUID:93075640
A;Accession: A49015
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-390,'Q',392-425,427-452 <WAT>
A;Cross-references: GB:M99437; NID:g182188; PID:g182189
A;Note: sequence extracted from NCBI backbone (NCBIN:119390, NCBIP:119392)
R;Baud, V.; Lipinski, M.; Rassart, E.; Poliquin, L.; Bergeron, D.
Genomics 11, 223-224, 1991
A;Title: The human homolog of the mouse common viral integration region, FLI1, maps t
A;Reference number: I54170; MUID:92112219
A;Accession: I54170
A;Status: preliminary; nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6 <RE2>
A;Cross-references: GB:L47616; NID:g1000863; PID:g1000864
C;Genetics:
A;Gene: GDB:FLI1
A;Cross-references: GDB:127565; OMIM:193067
A;Map position: 11q24.1-11q24.3
A;Introns: 77/3
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C;Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription
F;118-192/Domain: ets RII regulatory region homology <ETS2>
F;283-361/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 330 MNYDKLSR 337

RESULT 13
S29844
transforming protein fli, short splice form - human
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Homo sapiens (man)
C;Date: 02-Dec-1993 #sequence_revision 23-May-1997 #text_change 02-Sep-1997
C;Accession: S29844
R;Bromas, R.; May, W.; Denny, C.; Paskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Klems
Biochim. Biophys. Acta 1172, 155-158, 1993
A;Title: Human FLI-1 localizes to chromosome 11Q24 and has an aberrant transcript in
A;Reference number: S29843; MUID:93176799
A;Accession: S29844
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-385 <HRO>
A;Cross-references: EMBL:M93255; NID:g182655; PID:g182661
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1992
C;Genetics:
A;Gene: GDB:FLI1
A;Cross-references: GDB:127565, OMIM:193067
A;Map position: 11q24.1-11q24.3
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C;Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription
F;52-126/Domain: ets RII regulatory region homology <ETS2>
F;217-295/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 385;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 264 MNYDKLSR 271

RESULT 14
S17403
transforming protein fli - mouse
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Sep-1997
C;Accession: S17403

R;Bo[illegible], Y.; Giddens, E.B.; Letwin, K.; Bernstein, A.
Genes Dev. 5, 908-918, 1991
A;Title: Erythroleukemia induction by Friend murine leukemia virus: insertional activati
A;Reference number: S17403; MUID:91257578
A;Accession: S17403
A;Molecule type: mRNA
A;Residues: 1-452 <BEN>
A;Cross-references: GB:X53431; NID:g50974; PID:g50975
C;Genetics:
A;Gene: fli
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transforming pro
F;1-192/Domain: ets RII regulatory region homology <ETS2>
F;282-361/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 MNYIKLSR 367
||||||||
Db 357 MNYIKLSR 337

RESULT 15
S49013
transforming protein fli - African clawed frog
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Xenopus laevis (African clawed frog)
C;Date: 07-May-1995 #sequence_revision 23-May-1997 #text_change 05-Sep-1997
C;Accession: S49013
R;Meyer, D.; Wolff, C.M.; Stiegler, P.; Senan, F.; Befort, N.; Befort, J.J.; Remy, P.
Mech. Dev. 44, 109-121, 1993
A;Title: X. fli, the Xenopus homologue of the fli-1 gene, is expressed during embryogene
A;Reference number: S49013; MUID:94206844
A;Accession: S49013
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-453 <MEY>
A;Cross-references: EMBL:X66979; NID:g505486; PID:g505487
C;Genetics:
A;Gene: fli
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transforming pro
F;1[illegible]/Domain: ets RII regulatory region homology <ETS2>
F;2[illegible]/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 453;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 MNYIKLSR 367
||||||||
Db 358 MNYIKLSR 338

Search completed: November 20, 1999, 18.07:38
Job time: 23 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 18:07:15 ; Search time 11.23 Seconds
(without alignments)
1195.186 Million cell updates/sec

Title: US-09-126-945-2
Perfect score: 335
Sequence: 1 MSSASPGLSSVSPSHLLLPF.....GIIREEDISQRLVYQFVHPI 335

Scoring table: OLIGO

Searched: 122810 seqs, 40065486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	% Query Match	Length	DB	ID	Description
1	24	7.2	114	2	S28821	transcription fact
2	14	4.2	477	1	I38893	transcription fact
3	14	4.2	477	1	B46396	transcription fact
4	14	4.2	510	1	S43692	transcription fact
5	14	4.2	462	1	S35534	adenovirus E1A enh
6	14	4.2	555	1	S24061	transcription fact
7	10	3.0	453	2	B42093	serum response fac
8	10	3.0	430	2	I48755	mSAP1a - mouse
9	10	3.0	405	2	A42093	serum response fac
10	8	2.4	486	1	TVHUEG	transforming prote
11	8	2.4	478	1	S60754	transcription fact
12	8	2.4	452	1	I37565	transforming prote
13	8	2.4	385	1	S29844	transforming prote
14	8	2.4	452	1	S17403	transforming prote
15	8	2.4	453	1	S49013	transforming prote
16	8	2.4	428	1	TVHUEK	transforming prote
17	8	2.4	103	2	S28823	transcription fact
18	8	2.4	94	2	S28820	transcription fact
19	8	2.4	173	2	A56646	transcription fact
20	8	2.4	272	2	A54617	transcription fact
21	8	2.4	179	2	S51229	transcription fact
22	8	2.4	429	2	JC4965	elk1 protein - mou
23	8	2.4	409	2	I48680	transcription fact
24	8	2.4	407	2	I38062	Net - human
25	8	2.4	247	2	S72775	probable glucose-6
26	8	2.4	829	2	A34692	ecdysone-induced p
27	8	2.4	883	2	B34692	ecdysone-induced p
28	8	2.4	807	2	A53225	ecdysone-induced p
29	8	2.4	873	2	B53225	ecdysone-induced p
30	8	2.4	548	2	S59133	ETS2 repressor fac
31	8	2.4	250	2	A54308	PE-1 protein - hum
32	8	2.4	472	2	E71189	hypothetical prote
33	7	2.1	421	1	DEHUCM	acyl-CoA dehydroge
34	7	2.1	421	1	DERTCM	acyl-CoA dehydroge
35	7	2.1	444	1	B41864	acetyl-CoA:deacet
36	7	2.1	1748	1	JQ1555	genome polyprotein
37	7	2.1	2077	1	WZBE24	240K tegument prot
38	7	2.1	421	2	A55724	acyl-CoA dehydroge
39	7	2.1	270	2	S47099	site-specific DNA-

40	7	2.1	461	2	A59984	endo-1,4-beta-gluc
41	7	2.1	525	2	H70925	probable livD prot
42	7	2.1	686	2	A44842	cGMP-gated ion cha
43	7	2.1	133	2	I37382	cGMP-gated cation
44	7	2.1	521	2	S54266	glycoprotein gC -
45	7	2.1	411	2	A49127	homeotic protein A

ALIGNMENTS

RESULT 1
S28821
transcription factor ets-98B - fruit fly (Drosophila melanogaster) (fragment)
N;Alternate names: transforming protein ets-4
C;Species: Drosophila melanogaster
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 24-Sep-1998
C;Accession: S28821
R;Chen, L.; Bunting, M.; Karim, F.D.; Thummel, C.S.
Dev. Biol. 151, 176-191, 1992
A;Title: Isolation and characterization of five Drosophila genes that encode an ets-rela
A;Reference number: S28819; MUID:92249640
A;Accession: S28821
A;Molecule type: DNA
A;Residues: 1-114 <CHE>
A;Cross-references: EMBL:M88474; NID:g157195; PID:g157196
C;Genetics:
A;Gene: ET98B
A;Cross-references: FlyBase:FBgn0005659
C;Superfamily: ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;2[illegible] ets DNA-binding domain homology <ETS>

Query Match 7.2%; Score 24; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 3.2e-17;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 294 RKNRPAMNYDKLSRSIRQYYKKGI 317
||||||||||||||||||||||||
Db 71 RKNRPAMNYDKLSRSIRQYYKKGI 94

RESULT 2
I48893
transcription factor ERM - human
N;Alternate names: ets translocation variant 1
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 23-May-1997 #text_change 18-Sep-1998
C;Accession: I48893; S61541
R;Jeon, I.; Davis, J.N.; Braun, B.S.; Sublett, J.E.; Roussel, M.F.; Denny, C.T.; Shapi
Oncogene 10, 1229-1234, 1995
A;Title: A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS ge
A;Reference number: I48893; MUID:95215084
A;Accession: I48893
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-477 <RES>
A;Cross-references: EMBL:U17163; NID:g596005; PID:g596006
R;Monte, D.; Coutte, L.; Baert, J.L.; Angeli, I.; Stehelin, D.; de Launoit, Y.
Oncogene 11, 771-779, 1995
A;Title: Molecular characterization of the ets-related human transcription factor ER81.
A;Reference number: S61541; MUID:95380185
A;Accession: S61541
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-86,'V',40-60,'V',80-116,'C',118-126,'K',128-252,254-348,'A',350-477 <MON>
A;Cross-references: EMBL:X87175; NID:g1045060; PID:g1045061
C;Genetics:
A;Gene: GDB:ETV1
A;Cross-references: GDB:335222; OMIM:600541
A;Map position: 7q21.3-q22.3
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;337-415/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 477;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 379 KNRPAMNYDKLSRS 392

RESULT 3
B46396
transcription factor ER81 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-May-1994 #sequence_revision 23-May-1997 #text_change 05-Sep-1997
C;Accession: B46396
R;Brown, T.A.; McKnight, S.L.
Genes Dev. 6, 2502-2512, 1992
A;Title: Specificities of protein-protein and protein-DNA interaction of GABPalpha an
A;Reference number: A46396
A;Accession: B46396
A;Molecule type: mRNA
A;Residues: 1-477 <BRO>
A;Cross-references: GB:L10426; NID:g193195; PID:g515964
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;337-415/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 477;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 379 KNRPAMNYDKLSRS 392

RESULT 4
S43692
transcription factor erm - human
N;Alternate names: ets variant gene 5; ets-related molecule
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 23-May-1997 #text_change 18-Sep-1998
C;Accession: S43692
R;Monte, D.; Baert, J.L.; Defossez, P.A.; de Launoit, Y.; Stehelin, D.
Oncogene 9, 1397-1406, 1994
A;Title: Molecular cloning and characterization of human ERM, a new member of the Ets
A;Reference number: S43692; MUID:94203669
A;Accession: S43692
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-510 <MON>
A;Cross-references: EMBL:X76184; NID:g479166; PID:g479167
C;Genetics:
A;Gene: GDB:ETV5; ERM; ETV2L1
A;Cross-references: GDB:361087; OMIM:601600
A;Map position: 3q27-3q29
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;370-448/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 412 KNRPAMNYDKLSRS 425

RESULT 5
S35534
adenovirus E1A enhancer-binding protein E1AF - human (fragment)
N;Alternate names: ets translocation variant 4
C;Species: Homo sapiens (man)
C;Date: [illegible] 1993 #sequence_revision 23-May-1997 #text_change 18-Sep-1998
C;Accession: S35534
R;Higashino, F.; Yoshida, K.; Fujinaga, Y.; Kamio, K.; Fujinaga, K.
Nucleic Acids Res. 21, 547-553, 1993
A;Title: Isolation of a cDNA encoding the adenovirus E1A enhancer binding protein: a new
A;Reference number: S35534; MUID:93181246
A;Accession: S35534
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-462 <HIG>
A;Cross-references: GB:D12765; NID:g219610; PID:d1002729; PID:g219611
C;Genetics:
A;Gene: GDB:ETV4; E1A-F
A;Cross-references: GDB:512230; OMIM:600711
A;Map position: 17q21-17q21
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;321-[illegible]/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 462;
Best Local Similarity 100.0%; Pred. No. 2.6e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 363 KNRPAMNYDKLSRS 376

RESULT 6
S24061
transcription factor PEA3 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Feb-1994 #sequence_revision 23-May-1997 #text_change 05-Sep-1997
C;Accession: S24061
R;Xin, J.H.; Cowie, A.; Lachance, P.; Hassell, J.A.
Genes Dev. 6, 481-496, 1992
A;Title: Molecular cloning and characterization of PEA3, a new member of the Ets oncogen
A;Reference number: S24061; MUID:92192459
A;Accession: S24061
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-555 <XIN>
A;Cross-references: EMBL:X63190; NID:g53627; PID:g53628
C;Superfamily: transcription factor ER81; ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;414-472/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.2%; Score 14; DB 1; Length 555;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 KNRPAMNYDKLSRS 308
||||||||||||||
Db 456 KNRPAMNYDKLSRS 469

RESULT 7
B42093
serum response factor accessory protein-1 form a (SAP-1a) - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Aug-1998
C;Accession: B42093; S28917
R;Dalton, S.; Treisman, R.
Cell 68, 597-612, 1992
A;Title: Characterization of SAP-1, a protein recruited by serum response factor to t
A;Reference number: A42093; MUID:92154673
A;Accession: B42093
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-453 <DAL>
A;Cross-references: EMBL:M85165
A;Experimental source: HeLa cells
A;Note: sequence extracted from NCBI backbone (NCBIN:82522, NCBIP:82524)
C;Superfamily: elk-1 transforming protein; ets DNA-binding domain homology
C;Keywords: DNA binding
F;7-85/Domain: ets DNA-binding domain homology <ETS>

Query Match 3.0%; Score 10; DB 2; Length 453;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 VARLWGIRKN 296
||||||||||
Db 41 VARLWGIRKN 50

RESULT 8
I48755
mSAP1a - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Aug-1998
C;Accession: I48755
R;Giovane, A.; Pintzas, A.; Maira, S.M.; Sobieszczuk, P.; Wasylyk, B.
Genes Dev. 8, 1502-1513, 1994
A;Title: Net, a new ets transcription factor that is activated by Ras.
A;Reference number: A53837; MUID:95047310
A;Accession: I48755
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-430 <RES>
A;Cross-references: EMBL:Z36885; NID:g535127; PID:g535128
C;Superfamily: elk-1 transforming protein; ets DNA-binding domain homology
F;7-85/Domain: ets DNA-binding domain homology <ETS>

Query Match 3.0%; Score 10; DB 2; Length 430;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 VARLWGIRKN 296
||||||||||
Db 41 VARLWGIRKN 50

RESULT 9
A42093
serum response factor accessory protein-1 form b (SAP-1b) - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Aug-1998
C;Accession: A42093; S28816
R;Dalton, S.; Treisman, R.
Cell 68, 597-612, 1992
A;Title: Characterization of SAP-1, a protein recruited by serum response factor to t
A;Reference number: A42093; MUID:92154673
A;Accession: A42093
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-405 <DAL>
A;Cross-references: GB:M85164; NID:g338034; PID:g338035
A;Experimental source: HeLa cells
A;Note: sequence extracted from NCBI backbone (NCBIN:82525, NCBIP:82527)
C;Superfamily: elk-1 transforming protein; ets DNA-binding domain homology
C;Keywords: DNA binding
F;7-85/Domain: ets DNA-binding domain homology <ETS>

Query Match 3.9%; Score 10; DB 2; Length 405;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 VAPLWTIPKN 296
||||||||||
Db 41 VAPLWTIPKN 50

RESULT 9
TVHUE3
transforming protein erg-3 - human
N;Contains: transforming protein erg-1; transforming protein erg-2
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1989 #sequence_revision 30-May-1997 #text_change 31-Oct-1997
C;Accession: A94294; A94178; I58410; A28041; A29515
R;Rao, V.N.; Papas, T.S.; Shyam, E.; Reddy, P.
Science 237, 635-639, 1987
A;Title: erg, a human ets-related gene on chromosome 21: alternative splicing, polyadeny
A;Reference number: A94294; MUID:87263429
A;Accession: A94294
A;Molecule type: mRNA
A;Residues: 1-231,256-486 <REA>
A;Cross-references: GB:M17254; NID:g182186; PID:g182187
R;Reddy, E.S.P.; Rao, V.N.; Papas, T.S.
Proc. Natl. Acad. Sci. U.S.A. 84, 6131-6135, 1987
A;Title: The erg gene: a human gene related to the ets oncogene.
A;Reference number: A94178; MUID:87317698
A;Accession: A94178
A;Molecule type: mRNA
A;Residues: 100-231,256-486 <REE>
A;Cross-references: GB:M21535; NID:g182182; PID:g182185
R;Prasad, D.D.; Rao, V.N.; Lee, L.; Reddy, E.S.
Oncogene 5, 669-673, 1990
A;Title: Differentially spliced erg-3 product functions as a transcriptional activator
A;Reference number: I58410; MUID:94119611
A;Accession: I58410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 230-259 <RES>
A;Cross-references: GB:S68130; NID:g544960; PID:g544961
C;Genetics:
A;Gene: GDB:ERG
A;Cross-references: GDB:119884; OMIM:165080
A;Map position: 21q22.2-21q22.2
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription fa
F;1-231,256-486/Product: transforming protein erg-3 #status predicted <EG3>
F;1-231,256-486/Product: transforming protein erg-2 #status predicted <EG2>
F;1-231,256-486/Product: transforming protein erg-1 #status predicted <EG1>
F;126-207/Domain: ets RII regulatory region homology <ETS2>
F;320-398/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 467 MNYDKLSR 474

RESULT 11
S60754
transcription factor erg - chicken
C;Species: Gallus gallus (chicken)
C;Date: 27-Apr-1996 #sequence_revision 23-May-1997 #text_change 10-Jul-1998
C;Accession: S60754
R;Dhordain, P.; Dewitte, F.; Desbiens, X.; Stehelin, D.; Duterque-Coquillaud, M.
Mech. Dev. 50, 17-28, 1995
A;Title: Mesodermal expression of the chicken erg gene associated with precartilaginous
A;Reference number: S60754; MUID:95229425
A;Accession: S60754
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-478 <DHO>
A;Cross-references: EMBL:X77159; NID:g790439; PID:g790440
C;Genetics:
A;Gene: erg
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C;Keywords: DNA binding; phosphoprotein; proto-oncogene; transcription factor; transf
F;119-193/Domain: ets RII regulatory region homology <ETS2>
F;312-390/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 359 MNYDKLSR 366

RESULT 12
I37565
transforming protein fli, long splice form - human
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1996 #sequence_revision 30-May-1997 #text_change 20-Mar-1998
C;Accession: I37565; S29843; S35506; A49000; A49015; I54170
R;Delattre, O.; Zucman, J.; Plougastel, B.; Desmaze, C.; Melot, T.; Peter, M.; Kovar,
Nature 3599, 162-165, 1992
A;Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocatio
A;Reference number: I37378; MUID:92396229
A;Accession: I37565
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-452 <RES>
A;Cross-references: EMBL:X67001; NID:g32529; PID:g32530
R;Hromas, R.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Klems
Biochim. Biophys. Acta 1172, 155-158, 1993
A;Title: Human FLI-1 localizes to chromosome 11Q24 and has an aberrant transcript in
A;Reference number: S29843; MUID:93176799
A;Accession: S29843
A;Molecule type: mRNA
A;Residues: 1-68,'V',70-129,'A',131-132,'V',134-322,'Q',324-425,427-452 <HRO1>
A;Cross-references: EMBL:M93255
A;Note: the authors translated the codon GTG for residue 69 as Glu
R;Hromas, R.A.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Kle
submitted to the EMBL Data Library, May 1993
A;Description: Human FLI-1, an ETS oncogene family member preferentially expressed in
A;Reference number: S35506
A;Accession: S35506
A;Molecule type: mRNA
A;Residues: 1-68,'V',70-76,78-129,'A',131-132,'V',134-322,'Q',324-425,427-452 <HRO2>
A;Cross-references: EMBL:M93255; NID:g182659; PID:g182660
R;Prasad, D.D.; Rao, V.N.; Reddy, E.S.
Cancer Res. 52, 5833-5837, 1992
A;Title: Structure and expression of human Fli-1 gene.
A;Reference number: A49000; MUID:93007976
A;Accession: A49000
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-322,'Q',324-452 <PPA>
A;Cross-references: GB:S45205; NID:g257353; PID:g257354
A;Note: sequence extracted from NCBI backbone (NCBIN:115336, NCBIP:115337)
R;Watson, D.K.; Smyth, F.E.; Thompson, D.M.; Cheng, J.Q.; Testa, J.R.; Papas, T.S.; S
Cell Growth Differ. 3, 705-713, 1992
A;Title: The EPGB/Fli-1 gene: isolation and characterization of a new member of the f
A;Reference number: A49015; MUID:93075640
A;Accession: A49015
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-390,'Q',392-425,427-452 <WAT>
A;Cross-references: GB:M98833; NID:g182188; PID:g182189

A;N[illegible] sequence retrieved from NCBI backbone (NCBIN:119390, NCBIP:119392)
R;Bou[illegible] V.; Lipinski, M.; Bassert, F.; Poliquin, L.; Bergeron, D.
Genomics 11, 223-224, 1991
A;Title: The human homolog of the mouse common viral integration region, FLI1, maps to 1
A;Reference number: I54170; MUID:92112219
A;Accession: I54170
A;Status: preliminary; nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6 <RE2>
A;Cross-references: GB:L47616; NID:g1000863; PID:g1000864
C;Genetics:
A;Gene: GDB:FLI1
A;Cross-references: GDB:127565; OMIM:193067
A;Map position: 11q24.1-11q24.3
A;Introns: 77/3
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription fa
F;118-192/Domain: ets RII regulatory region homology <ETS2>
F;282-361/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 330 MNYDKLSR 337

RESULT 13
S29844
transforming protein fli, short splice form - human
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Homo sapiens (man)
C;Date: 02-Dec-1993 #sequence_revision 23-May-1997 #text_change 02-Sep-1997
C;Accession: S29844
R;Hromas, R.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Klemsz,
Biochim. Biophys. Acta 1172, 155-158, 1993
A;Title: Human FLI-1 localizes to chromosome 11Q24 and has an aberrant transcript in neu
A;Reference number: S29843; MUID:93176799
A;Accession: S29844
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-385 <HRO>
A;Cross-references: EMBL:M93255; NID:g182655; PID:g182661
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1992
C;Genetics:
A;Gene: GDB:FLI1
A;Cross-references: GDB:127565; OMIM:193067
A;Map position: 11q24.1-11q24.3
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription fa
F;52-126/Domain: ets RII regulatory region homology <ETS2>
F;217-296/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 385;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 264 MNYDKLSR 271

RESULT 14
S17403
transforming protein fli - mouse
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Sep-1997
C;Accession: S17403
R;Ben-David, Y.; Giddens, E.B., Letwin, K., Bernstein, A.
Genes Dev. 5, 908-918, 1991
A;Title: Erythroleukemia induction by Friend murine leukemia virus: insertional activ
A;Reference number: S17403; MUID:91257578
A;Accession: S17403
A;Molecule type: mRNA
A;Residues: 1-452 <BEN>
A;Cross-references: GB:X59421; NID:g50974; PID:g50975
C;Genetics:
A;Gene: Fli
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C;Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transforming
F;118-192/Domain: ets RII regulatory region homology <ETS2>
F;283-361/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 330 MNYDKLSR 337

RESULT 15
S49013
transforming protein fli - African clawed frog
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Xenopus laevis (African clawed frog)
C;Date: 07-May-1995 #sequence_revision 23-May-1997 #text_change 05-Sep-1997
C;Accession: S49013
R;Meyer, D.; Wolff, C.M.; Stiegler, P.; Senan, F.; Befort, N.; Befort, J.J.; Remy, P.
Mech. Dev. 44, 109-121, 1993
A;Title: Xl-fli, the Xenopus homologue of the fli-1 gene, is expressed during embryog
A;Reference number: S49013; MUID:94206844
A;Accession: S49013
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-453 <MEY>
A;Cross-references: EMBL:X66979; NID:g505486; PID:g505487
C;Genetics:
A;Gene: fli
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C;Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transforming
F;117-191/Domain: ets RII regulatory region homology <ETS2>
F;284-362/Domain: ets DNA-binding domain homology <ETS>

Query Match 2.4%; Score 8; DB 1; Length 453;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 300 MNYDKLSR 307
||||||||
Db 331 MNYDKLSR 338

Search completed. November 20, 1999, 18:07.38
Job time: 23 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 1999, 15:16:47 ; Search time 505.23 Seconds
(without alignments)
11922.305 Million cell updates/sec

Title: US-09-126-945-1
Perfect score: 1894
Sequence: 1 atgtgacttcctcccaggcac ... .ataaagatactagaggaactg 1894

Scoring table: OLIGODX_NUC

Searched: 679419 seqs, 1590154680 residues

Database: GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_prl:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1894	100.0	1894	11	AF071538	AF071538 Homo sapi
2	38	2.0	1704	12	AB019436	AB019436 Mus muscu
3	30	1.6	848	4	AF075706	AF075706 Gallus ga
4	29	1.5	2508	4	AF075708	AF075708 Gallus ga
c 5	28	1.5	60056	35	AC007818	AC007818 Drosophil
6	28	1.5	362	36	DRODETS4A	M88474 Drosophila
7	25	1.3	1403	10	HSDNAFEV3	Y08979 H.sapiens F
8	25	1.3	1901	10	HSRNAFEV	Y08976 H.sapiens m
9	24	1.3	480	36	DRODETS9A	M88475 Drosophila
c 10	24	1.3	85095	37	AC004573	AC004573 Drosophil
c 11	22	1.2	115835	9	HS28C20	AL031119 Human DNA
12	22	1.2	2658	12	MMNETRN	Z32815 M.musculus
13	22	1.2	2211	12	MUSTRANCNC	L19953 Mus musculu
14	22	1.2	1526	12	S82864	S82864 Elk-3=Ets t
c 15	22	1.2	113899	36	PFMAL3P4	AL008970 Plasmodiu
16	21	1.1	131453	34	HS989H11	Z83851 Homo sapien
17	20	1.1	128598	1	D90912	D90912 Synechocyst
18	20	1.1	473	5	E06024	E06024 cDNA encodi
19	20	1.1	2072	5	E06025	E06025 cDNA encodi
c 20	20	1.1	36742	9	D87002	D87002 Human (lamb
21	20	1.1	2064	9	HUMELAF	D12765 Human mRNA
c 22	20	1.1	152141	10	HSU07000	U07000 Human break
23	20	1.1	2667	10	HSU15655	U15655 Human ets d
24	20	1.1	2333	10	HSU18018	U18018 Human E1A e
25	20	1.1	35687	11	AC000051	AC000051 Homo sapi
26	20	1.1	34539	11	AC002051	AC002051 Homo sapi
c 27	20	1.1	151111	11	AC002308	AC002308 Homo sapi
c 28	20	1.1	36921	11	AC004185	AC004185 Homo sapi
c 29	20	1.1	112220	11	AC005373	AC005373 Homo sapi
c 30	20	1.1	37139	11	AC006047	AC006047 Homo sapi
c 31	20	1.1	111084	11	AC006486	AC006486 Homo sapi
32	20	1.1	2642	11	HSETSF5	AF095890 Homo sapi
33	20	1.1	1046	12	MUSETSRPA	L10427 Mus musculu
c 34	20	1.1	126983	34	AC005283	AC005283 Homo sapi
c 35	20	1.1	236822	34	D84394	D84394 Homo sapien
c 36	20	1.1	131930	34	HSB7L1C4	AL078476 Homo sapi
37	20	1.1	269150	34	HSDJ94E24	AL050317 Homo sapi
c 38	20	1.1	160979	35	AC007664	AC007664 Homo sapi
c 39	20	1.1	102353	35	AC007708	AC007708 Homo sapi
40	20	1.1	58839	35	AC007821	AC007821 Drosophil
41	20	1.1	1832	37	AF106964	AF106964 Strongylo
42	20	1.1	43681	42	AC002054	AC002054 Homo sapi
c 43	20	1.1	155313	42	AC006459	AC006459 Homo sapi
44	19	1.0	35516	1	MTCY373	Z73419 Mycobacteri
45	19	1.0	40790	1	MTCY493	Z95844 Mycobacteri

ALIGNMENTS

RESULT 1
AF071538
LOCUS AF071538 1894 bp mRNA PRI 13-DEC-1998
DEFINITION Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete cds.
ACCESSION AF071538
NID g4007417
VERSION AF071538.1 GI:4007417
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1894)
AUTHORS Oettgen,P., Finger,E., Akbarali,Y., Thamrongsak,U., Dube,A., Boltax,J., Brown,L., Kunsch,C. and Libermann,T.A.
TITLE Isolation and characterization of a novel prostate epithelium-specific Ets transcription factor, PDEF
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1894)
AUTHORS Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical Center, 330 Brookline Ave., Boston, MA 02215, USA

```
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium-specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSPTGLEKAAAGAVGLERPDWS
                     PSPPATPEQGLSAPYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WIEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGAIHYCASLSEPSWIDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     CIEKIEDSAQVARLWGIRKNRPAMNYDKLSRSIPQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT       468 a    654 c    571 g    302 t
ORIGIN
```

Query Match 100.0%; Score 1894; DB 11; Length 1894;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atctgaacttcctcccagcacattcctgcactctgccgtgtccacactgccccacagaccc  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTCTGACTTCCTCCCAGCACATTCCTGCACTCTGCCGTGTCCACACTGCCCCACAGACCC  60

Qy      61 agttcctcagcctgctgtcagcctccctgcaagccctcaggttgggccttgccacggtg  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGTTCTCCAAGCCTGCTGGCAGCTCCCTGCAAGCCCCTCAGGTTGGGCCTTGCCACGGTG  120

Qy     121 ccagcagacagcctgagctgagggtagggactccctacaggcacgcagccctgagacc  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCAGCAGTCAGCCCTGGCCTGGGGTAGGGACTCCCTACAGGCACGCAGCCCTGAGACC  180

Qy     181 tcagcagcagcagccctgagctgtgccaggccccagtggccaacctgagtgctgcctct  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCAGAGGCCACCCCTTGAGGTGGCCAGGCCCCCAGTGGCCAACCTGAGTGCTGCCTCT  240

Qy     241 ccccccagcgccctgctgaccccctagttccgctggcccccccagatgcctggctgagacacgc  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCCACCAGCCCTGCTGGCCCCTGGTTCCGCTGGCCCCCCAGATGCCTGGCTGAGACACGC  300

Qy     301 cagtggcctcagctgcccacacctcttccccgccccctgaagttggcactgcagcagacag  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGTGGCCTCAGCTGCCCACACCTGTTCCCGGCCCCTGAAGTTGGCACTGCAGCAGACAG  360

Qy     361 ctcccctgagcagcccagcctaacagacacagccgccagcccaaacagcagcggcatggg  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTCCCTGGGCAGCAGGCAGCTAACAGACACAGCGGCCAGCCCAAACAGCAGCGGCATGGG  420

Qy     421 tcccccagcccccccctctgaacaagctgtatcccccagccacctcctgctgcccccccgacac  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CACCCCAGCCCCCCCTCTGAGCAGCGTATCCCCCAGCCACCTCCTGCTGCCCCCCGACAC  480

Qy     481 agtgtcgcggaacaagctttgaagaaggcggcagccagcggggcagtgggtctcgaagagacggga  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGTGTCGCGGAACAAGCTTTGAGAAGGCGGCAGCCGGGGCAGTGGGTCTCGAGAGACGGGA  540

Qy     541 ctgaagtccagtcaccccgcagccccgagcaggcctgtccgccttctacctctcctcta  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CTGGAGTCCAGTCCACCCCGCAGCCCCGAGCAGGGCCTGTCCGCCTTCTACCTCTCCTA  600

Qy     601 ctttgacatgctgtacctgaggacagcagccgggcagccaaggcccctgggggccagcag  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CTTTGACATGCTGTACCCTGAGGACAGCAGCCGGGCAGCCAAGGCCCCTGGGGGCCAGCAG  660

Qy     661 tcgggaaggagccacctgaggagcctgagcagtgccccgtccattgacagccaaqccccagc  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TCGGGAGGAGCCACCTGAGGAGCCTGAGCAGTGCCCGGTCATTGACAGCCAAGCCCAGC  720

Qy     721 gggcagcctggacttggtgcccggcgggctgaccttgaaggaggcactcgctggagccaggt  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGGCAGCCTGGACTTGGTGCCCGGCGGGCTGACCTTGGAGGAGCACTCGCTGGAGCAGGT  780

Qy     781 gcagtccatggtggtgggcgaagtgctcaaggacatcgaggcggcctgcaaggctgctcaa  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCAGTCCATGGTGGTGGGCGAAGTGCTCAAGGACATCGAGGCGGCCTGCAAGCTGCTCAA  840

Qy     841 catcaccgcagatcccatggactgtggagcccagcaatatgcaggaagtggctcctctgtggac  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CATCACCGCAGATCCCATGGACTGTGGAGCCCAGCAATATGCAGGAAGTGGCTCCTCTGTGGAC  900

Qy     901 agagcaccaatacccggctgccccccatgggcaaggccttccaggagctggcgggcaaggga  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGAGCACCAATACCGGCTGCCCCCCATGGGCAAGGCCTTCCAGGAGCTGGCGGGCAAGGGA  960

Qy     961 gctgtgcgccatgtcggaggagcagttccgccagcgctcgccccctgggtgggggatgtgct 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCTGTGCGCCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGGGATGTGCT 1020

Qy    1021 gcacgcccacctggacatctggaagtcagcggcctgggtgaaaggagcggacttcaccctgg 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGGTGAAAGAGCGGACTTCACCCTGG 1080

Qy    1081 ggcgattcactactgtgcctcgaccagtgaggagagctggaccgacagcgaggtgggactc 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GGCGATTCACTACTGTGCCTCGACCAGTGAGGAGAGCTGGACCGACAGCGAGGTGGGACTC 1140

Qy    1141 atcatgctccgggcagcccatccacctgtggcagttcctcaaggagttgctactcaagcc 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ATCATGCTCCGGGCAGCCCATCCACCTGTGGCAGTTCCTCAAGGAGTTGCTACTCAAGCC 1200

Qy    1201 ccacagctatggccgcttcattaggtggctcaacaaggaggaaggggcatcttcaaaattga 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CCACAGCTATGGCCGCTTCATTAGGTGGCTCAACAAGGAGGAAGGGGCATCTTCAAAATTGA 1260

Qy    1261 ggactcagcccaggtggcccggctgtggggcatccgcaagaacggtcccgccatgaactta 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GGACTCAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAAGAACGGTCCCGCCATGAACTA 1320

Qy    1321 cgacaagctgagccgctccatccgccagtattacaagaagggcatcatccggaagccaga 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCATCATCCGGAAGCCAGA 1380

Qy    1381 catctcccagcgcctcgtctaccagttcgtgcaccccatctgagtgcctggcccagggcc 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACCCCATCTGAGTGCCTGGCCCAGGGCC 1440

Qy    1441 tgaaacccgccctcagggccctctctcctgccctgccctgcctcagccagggccctgagatg 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TGAAACCCGGCCCTCAGGGCCCTCTCTCCTGCCCTGCCCTGCCTCAGCCAGGGCCTGAGATG 1500

Qy    1501 ggggaaaacgggcagtctgctctgctgctctgaccttccagagcccaagatcagggaggg 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GGGGAAAACGGGCAGTCTGCTCTGCTGCTCTGACCTTCCAGAGCCCAAGATCAGGGAGGG 1560

Qy    1561 gcaaccaactgcccccagggggatatgggtcctctgggccttcgggaccatgggccaggg 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GCAACCAACTGCCCCAGGGGGATATGGGTCCTCTGGGGCCTTCGGGACCATGGGCCAGGG 1620

Qy    1621 gtgcttcctcctcaggcccagctgctcccctggaggacagagggagacagggctgctccc 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GTGCTTCCTCCTCAGGCCCAGCTGCTCCCCTGGAGGACAGAGGGAGACAGGGCTGCTCCC 1680

Qy    1681 caacacctgcctctgaccccagcatttccagagcagagcctcacagaaggcagtgaatcag 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1741 CAACACCTGCCTCTGACCCCAGCATTTCCAGAGCAGAGCCTACAGAAGGGCAGTGACTCG 1740

Qy    1741 aaaaaaccacaaacagtccaaggcctctctctgctccatccccctgcctcccattctgca 1800
           ||| |||||| || | ||||||||||||||||||||||||||||||||||||||||||||
Db    1741 ATAAAGCCCACAGGCAGTCCAGGCCTCTCTCTGCTCCATCCCCCTGCCTCCCATTCTGCA 1800

Qy    1801 ccacacctggcatggtgcaggggagacatctgcacccctgagttgggcagccaggagtgcc 1860
           |||||| |||||| || | ||||||||||||||||||||||||||||||||||||||||||
Db    1801 CCACACCTGGCATGGTGCAGGGAGACATCTGCACCCCTGAGTTGGGCAGCCAGGAGTGCC 1860

Qy    1861 ccccggaatggataataaagatactagagaactg 1894
           |||||||| ||||||||||||||||||||||||||
Db    1861 CCCGGGAATGGATAATAAAGATACTAGAGAACTG 1894


RESULT   2
AB019436
LOCUS       AB019436     1704 bp    mRNA            ROD       08-MAY-1999
DEFINITION  Mus musculus mRNA for Prostate specific transcription factor ets,
            complete cds.
ACCESSION   AB019436
NID         g4760536
VERSION     AB019436.1  GI:4760536
KEYWORDS    Prostate specific transcription factor ets; mPse.
SOURCE      Mus musculus (strain:C57BL/6) Adult male Colon Epithelial cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Prostate specific ets
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1. .1704
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /cell_type="Epithelial"
                     /dev_stage="Adult"
                     /sex="male"
                     /tissue_type="Colon"
     gene            362. .1339
                     /gene="mPse"
     CDS             362. .1339
                     /gene="mPse"
                     /codon_start=1
                     /product="Prostate specific transcription factor ets"
                     /protein_id="BAA77329.1"
                     /db_xref="PID:d1041095"
                     /db_xref="PID:g4760537"
                     /db_xref="GI:4760537"
                     /translation="MGSASPGLSNVSPGCLLLFPDVAPRTGTEKAASGAMGPEKQEWS
                     PSPPAIPEQGLSAFYLSYFNMYPDDSSWVAKVPEARAGEDHPEEPEQCPVIDSQASGS
                     TLDEHSLEQVQSMVVGEVLKDIETACKLLNITADPGDWSPGNVQKWLLWTEHQYRLPP
                     AGKAFQFDVGKEICAMSEEQFPQPAPLGGDVLHAHLDIWKSAAWMKERTSPGTLHYCA
                     STSEEGWTDGFVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEKGIFKIEDSAQ
                     VAPTWGVRKNRPAMNYDKLSRSIPQYYKKGIIPKPDISQRLVYQFVHPV"
     polyA_site      1704
                     /note="10 a nucleotides"
BASE COUNT      369 a    527 c    491 g    317 t
ORIGIN


  Query Match             2.0%;  Score 38;  DB 12;  Length 1704;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-10;
  Matches   38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1025 gcccacctggacatctggaagtcagcggcctggatgaa 1062
          |||||||||||||||||||||||||||||||||||||
Db    941 GCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAA 978


RESULT   3
AF075706
LOCUS       AF075706      848 bp    mRNA            VRT       19-DEC-1998
DEFINITION  Gallus gallus ets domain protein (ERM) mRNA, partial cds.
ACCESSION   AF075706
NID         g3869357
VERSION     AF075706.1  GI:3869357
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 848)
  AUTHORS   Lin,J.H., Saito,T., Anderson,D.J., Lance-Jones,C., Jessell,T.M. and
            Arber,S.
  TITLE     Functionally related motor neuron pool and muscle sensory afferent
            subtypes defined by coordinate ETS gene expression
  JOURNAL   Cell 95 (3), 393-407 (1998)
  MEDLINE   99029846
REFERENCE   2  (bases 1 to 848)
  AUTHORS   Lin,J.H., Saito,T., Anderson,D.J., Lance-Jones,C., Jessell,T.M. and
            Arber,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1998) Center for Neurobiology & Behavior,
            Columbia University, 701 W. 168th St. HHSC 1013, New York, NY
            10032, USA
FEATURES             Location/Qualifiers
     source          1. .848
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            <1. .848
                     /gene="ERM"
     CDS             <1. .712
                     /gene="ERM"
                     /codon_start=2
                     /product="ets domain protein"
                     /protein_id="AAC97201.1"
                     /db_xref="PID:g3869358"
                     /db_xref="GI:3869358"
                     /translation="RPQPARPPPRHGFQSPMGIKQEPRDYCIDSEVPNPQSSYVRGGLF
                     PSSHDGFSYEKDTFLYFDDTCVVPERLEGKVKQEPILYPEGPPYQRRGSLQLWQFLVT
                     LLDDPANAHFIAWTGRGMEFKLIEPEEVARPWGIQKNRPAMNYDKLSRSLRYYYEKGI
                     MQKVAGERYVYKFVCDPDALFSMAYPDNQRPFLKAEPDCHVSEEDTLPLTHFEDSPAY
                     LLEVEHCGSLPYAEGFAY"
BASE COUNT      192 a    265 c    245 g    146 t
ORIGIN


  Query Match             1.6%;  Score 30;  DB 4;  Length 848;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-05;
  Matches   30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1310 gccatgaactacgacaagctgagccgctcc 1339
          ||||||||||||||||||||||||||||||
Db    425 GCCATGAACTACGACAAGCTGAGCCGCTCC 454


RESULT   4
AF075708
LOCUS       AF075708     2508 bp    mRNA            VRT       19-DEC-1998
DEFINITION  Gallus gallus ets domain protein (PEA3) mRNA, complete cds.
ACCESSION   AF075708
NID         g4033764
VERSION     AF075708.1  GI:4033764
KEYWORDS    .
```

```
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2508)
  AUTHORS   Lin,J.H., Saito,T., Anderson,D.J., Lance-Jones,C., Jessell,T.M. and
            Arber,S.
  TITLE     Functionally related motor neuron pool and muscle sensory afferent
            subtypes defined by coordinate ETS gene expression
  JOURNAL   Cell 95 (4), 393-407 (1998)
  MEDLINE   99020946
REFERENCE   2  (bases 1 to 2508)
  AUTHORS   Lin,J.H., Saito,T., Anderson,D.J., Lance-Jones,C., Jessell,T.M. and
            Arber,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1998) Center for Neurobiology & Behavior,
            Columbia University, 701 W. 168th St. HHSC 1013, New York, NY
            10032, USA
COMMENT     On Dec 19, 1998 this sequence version replaced gi:3869361.
FEATURES             Location/qualifiers
     source          1..2508
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            1..2508
                     /gene="PEA3"
     CDS             1..1011
                     /gene="PEA3"
                     /codon_start=1
                     /product="ets domain protein"
                     /protein_id="AAC97203.1"
                     /db_xref="PID:g4033765"
                     /db_xref="GI:4033765"
                     /translation="VHNNHRIVILQYLFDSPSAYDQPRPVGIKSPDPGIPIQSPFQQF
                     PRQDRNFLTPSGPPHPTSSWGYLIRNSSAYQAPVEMCHSFHSSQGMAQDAPIVYQRPM
                     SEQCLQCPPQSFKQEFHDPQYEQAGRSCQYPAAVVIKQEQVDYMYDSGCSYDKQMRPF
                     ADDVCVVPEKFEGDIKQEVGQYPEGPPYQPPGSLQLWQFLVALLDDPTNSHFIAWTGR
                     GMEFKLIEPEEVARLWGIQKNRPAMNYDKLSRSLRYYYEKGIMQKVAGERYVYKFVCE
                     PEALFSLAFPDNQRPALKAELEPQISEEDTVPLSHLDESTAYLPDLGSLPPQLYGKGY
                     IY"
BASE COUNT      565 a    715 c    672 g    557 t      1 others
ORIGIN


  Query Match          1.7%;  Score 29;  DB 4;  Length 2508;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-05;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1331 gcccatgaactacgacaaagctgagccgctc 1359
             |||||||||||||||||||||||||||||||
Db       724 GCCCATGAACTACGACAAGCTGAGCCGCTC 752


RESULT  5
AC007818/c
LOCUS       AC007818     60056 bp    DNA             HTG       14-JUN-1999
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR02M05 (D785) RPCI-98
            02.M.5 map 98B-98B strain y; cn bw sp, WORKING DRAFT SEQUENCE, 20
            unordered pieces.
ACCESSION   AC007818
NID         g5054146
VERSION     AC007818.1  GI:5054146
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 60056)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60056)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 400 bases, phrap computed error
            rate <= 1/10.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      893: contig of 893 bp in length
            *      894      973: gap of unknown length
            *      974     1964: contig of 991 bp in length
            *     1965     2044: gap of unknown length
            *     2045     3652: contig of 1608 bp in length
            *     3653     3732: gap of unknown length
            *     3733     5272: contig of 1540 bp in length
            *     5273     5352: gap of unknown length
            *     5353     6476: contig of 1124 bp in length
            *     6477     6556: gap of unknown length
            *     6557     8669: contig of 2113 bp in length
            *     8670     8749: gap of unknown length
            *     8750    11281: contig of 2532 bp in length
            *    11282    11361: gap of unknown length
            *    11362    13275: contig of 1914 bp in length
            *    13276    13355: gap of unknown length
            *    13356    16154: contig of 2799 bp in length
            *    16155    16234: gap of unknown length
            *    16235    18409: contig of 2175 bp in length
            *    18410    18489: gap of unknown length
            *    18490    20047: contig of 1558 bp in length
            *    20048    20127: gap of unknown length
            *    20128    21987: contig of 1860 bp in length
            *    21988    22067: gap of unknown length
            *    22068    24914: contig of 2847 bp in length
            *    24915    24994: gap of unknown length
            *    24995    28786: contig of 3792 bp in length
            *    28787    28866: gap of unknown length
            *    28867    32920: contig of 4054 bp in length
            *    32921    33000: gap of unknown length
            *    33001    36522: contig of 3522 bp in length
            *    36523    36602: gap of unknown length
            *    36603    41889: contig of 5287 bp in length
            *    41890    41969: gap of unknown length
            *    41970    46511: contig of 4542 bp in length
            *    46512    46591: gap of unknown length
            *    46592    51474: contig of 4883 bp in length
            *    51475    51554: gap of unknown length
            *    51555    60056: contig of 8502 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1. .60056
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /clone="BACR02M05 (D785) RPCI-98 02 M.5"
                     /chromosome="3"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6"
                     /map="98B-98B"
BASE COUNT    15556 a  13397 c  13635 g  15945 t   1523 others
ORIGIN


  Query Match             1.5%;  Score 28;  DB 35;  Length 60056;
  Best Local Similarity   100.0%;  Pred. No. 0.00018;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1159 catgcaacctgtggcagttcctcaaggag 1186
           ||||||||||||||||||||||||||||||
Db   54019 CATGCAACCTGTGGCAGTTCCTCAAGGAG 54048


RESULT   6
DROETS4A
LOCUS       DROETS4A     362 bp    DNA             INV       05-AUG-1992
DEFINITION  Drosophila melanogaster D-ets-4 DNA binding domain protein gene, 3'
            end.
ACCESSION   M88474
NID         g157195
VERSION     M88474.1  GI:157195
KEYWORDS    DNA-binding domain protein
SOURCE      Drosophila melanogaster DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Chen,T., Bunting,M., Karim,F.D. and Thummel,C.S.
  TITLE     Isolation and characterization of five Drosophila genes that encode
            an ets-related DNA binding domain
  JOURNAL   Dev. Biol. 151, 176-191 (1992)
  MEDLINE   92249640
FEATURES             Location/Qualifiers
     source          1. .362
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="Chromosome 3R 98B"
     gene            1. .345
                     /gene="Ets98B"
                     /note="Ets at 98B"
                     /allele=""
                     /db_xref="FlyBase:FBgn0005659"
     CDS             1. .345
                     /partial
                     /gene="Ets98B"
                     /codon_start=1
                     /product="D-ets-4 DNA binding domain protein"
                     /protein_id="AAA28451.1"
                     /db_xref="PID:g157196"
                     /db_xref="GI:157196"
                     /translation="TNASNGGIAIVKRPNGGRTGGGGSHIHLWQFLKELLASPQVNGT
                     AIRWIDPSKGIFKIEDSVPVAKLWGRRKNRPAMNYDKLSRSIRQYYKKGIMKKTERSQ
                     RLVYQFCHPYSQ"
BASE COUNT       90 a     99 c    113 g     60 t
ORIGIN      Chromosome 3R, 98B.


  Query Match             1.5%;  Score 28;  DB 36;  Length 362;
  Best Local Similarity   100.0%;  Pred. No. 0.00017;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1159 catgcaacctgtggcagttcctcaaggag 1186
           ||||||||||||||||||||||||||||||
Db      75 CATGCAACCTGTGGCAGTTCCTCAAGGAG 102


RESULT   7
HSDNAFEV3
LOCUS       HSDNAFEV3    1403 bp    DNA             PRI       23-APR-1997
DEFINITION  H.sapiens FEV gene, exon 3.
ACCESSION   Y08979
NID         g1841699
VERSION     Y08979.1  GI:1841699
KEYWORDS    ets gene family.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1403)
  AUTHORS   Peter,M., Couturier,J., Pacquement,H., Michon,J., Thomas,G.,
            Magdelenat,H. and Delattre,O.
  TITLE     A new member of the ETS family fused to EWS in Ewing tumors
  JOURNAL   Oncogene 14 (10), 1159-1164 (1997)
  MEDLINE   97236656
REFERENCE   2  (bases 1 to 1403)
  AUTHORS   Delattre,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-1996) O. Delattre, Institut Curie, Inserm U434,
            26 Rue Dulm, Paris 75231 Cedex 05, Paris 75231 Cedex 05, FRANCE
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 1403)
  AUTHORS   Delattre,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-1997) O. Delattre, Institut Curie, Inserm U434,
            26 Rue Dulm, Paris 75231 Cedex 05, Paris 75231 Cedex 05, FRANCE
COMMENT     On Feb 15, 1997 this sequence version replaced gi:1834485.
FEATURES             Location/Qualifiers
     source          1. .1403
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /germline
     intron          1. .240
                     /number=2
                     /evidence=experimental
     gene            241. .1403
                     /gene="FEV"
     exon            241. .1403
                     /gene="FEV"
                     /note="ETS family member"
                     /number=3
                     /evidence=experimental
BASE COUNT      210 a    551 c    376 g    266 t
ORIGIN


  Query Match             1.3%;  Score 25;  DB 10;  Length 1403;
  Best Local Similarity   100.0%;  Pred. No. 0.0083;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1312 catgaactacgacaagctgagccgc 1336
           |||||||||||||||||||||||||
Db     398 CATGAACTACGACAAGCTGAGCCGC 422


RESULT   8
HSRNAFEV
LOCUS       HSRNAFEV     1901 bp    mRNA            PRI       23-APR-1997
DEFINITION  H.sapiens mRNA for FEV protein.
ACCESSION   Y08976
NID         g1834508
VERSION     Y08976.1  GI:1834508
KEYWORDS    ets gene family.
SOURCE      human
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
```

```
           Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 1901)
  AUTHORS  Peter,M., Couturier,J., Pacquement,H., Michon,J., Thomas,G.,
           Magdelenat,H. and Delattre,O.
  TITLE    A new member of the ETS family fused to EWS in Ewing tumors
  JOURNAL  Oncogene 14 (10), 1159-1164 (1997)
  MEDLINE  97236656
REFERENCE  2  (bases 1 to 1901)
  AUTHORS  Delattre,O.
  TITLE    Direct Submission
  JOURNAL  Submitted (10-OCT-1996) O. Delattre, Institut Curie, Inserm U434,
           26 Rue Dulm, Paris 75231 Cedex 05, Paris 75231 Cedex 05, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1901
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /cell_line="ewing"
     CDS             585. .1301
                     /function="DNA binding protein"
                     /note="ETS family member"
                     /codon_start=1
                     /product="FEV protein"
                     /protein_id="CAA70169.1"
                     /db_xref="PID:e291921"
                     /db_xref="PID:g1834509"
                     /db_xref="GI:1834509"
                     /db_xref="SPTREMBL:Q99581"
                     /translation="MRQSGASQPLLINMYLPDPVGDGLFKDGKNPSWGPLSPAVQKGS
                     GQIQLWQFLLELLADRANAGCIAWEGGHGEFKLTDPDEVARRWGERKSKPNMNYDKLS
                     RALRYYYDKNIMSKVHGKRYAYRFDFQGLAQACQPPPAHAHAAAAAAAAAAAAQDGAL
                     YKLPAGLAPLPFPGLSKLNLMAASAGVAPAGFSYWPGPGPAATAAAATAALYPSPSLQ
                     PPPGPEGAVAAASHLGGHYH"
     misc_feature    711. .962
                     /note="DNA binding domain"
BASE COUNT      402 a    707 c    575 g    317 t
ORIGIN


  Query Match             1.3%;  Score 25;  DB 10;  Length 1901;
  Best Local Similarity   100.0%;  Pred. No. 0.0083;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1312 catgaactacgacaagctgagccg 1336
           |||||||||||||||||||||||||
Db     256 CATGAACTACGACAAGCTGAGCCG 280


RESULT  9
D00834/c
LOCUS       D00834S6A     480 bp    DNA             INV       05-AUG-1992
DEFINITION  Drosophila melanogaster D-ets-6 DNA binding domain protein gene,
            partial cds.
ACCESSION   M00847
NID         g157197
VERSION     M00847.1  GI:157197
KEYWORDS    DNA-binding domain protein
SOURCE      Drosophila melanogaster DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila
REFERENCE  1  (bases 1 to 480)
  AUTHORS  Chen,I., Bunting,M., Karim,F.D. and Thummel,C.S.
  TITLE    Isolation and characterization of five Drosophila genes that encode
           an ets-related DNA binding domain
  JOURNAL  Dev. Biol. 151, 176-191 (1992)
  MEDLINE  92249640
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="Chromsome 2L 21C "
     intron          1. .95
                     /gene="D-ets-6 DNA-binding domain"
     exon            96. .288
                     /gene="Ets21C"
                     /db_xref="FlyBase:FBgn0005660"
     gene            join(96. .288,366. .480)
                     /gene="Ets21C"
                     /note="Ets at 21C"
                     /allele=""
                     /db_xref="FlyBase:FBgn0005660"
     CDS             join(96. .288,366. .480)
                     /partial
                     /gene="Ets21C"
                     /codon_start=3
                     /product="D-ets-6 DNA binding domain protein"
                     /protein_id="AAA28452.1"
                     /db_xref="PID:g157198"
                     /db_xref="GI:157198"
                     /translation="SGGQIQLWQFLLELLADSSNANAISWEGQSGEFRLIDPDEVARR
                     WGEPKAKPNMNYDKLSPALRYYYDKNIMTKVHGKRYAYKFDFHGLMAACQAQAQGGDP
                     "
     intron          289. .365
                     /gene="D-ets-6 DNA-binding domain"
     exon            366. .480
                     /gene="Ets21C"
                     /db_xref="FlyBase:FBgn0005660"
BASE COUNT      101 a    138 c    142 g     99 t
ORIGIN      Chromosome 2L, 21C.


  Query Match             1.3%;  Score 24;  DB 36;  Length 480;
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1312 catgaactacgacaagctgagccg 1335
           |||||||||||||||||||||||||
Db     256 CATGAACTACGACAAGCTGAGCCG 279


RESULT  10
AC004573/c
LOCUS       AC004573    85095 bp    DNA             INV       30-JAN-1999
DEFINITION  Drosophila melanogaster, chromosome 2L, region 21C5-21D1, PI clone
            DS07610, complete sequence.
ACCESSION   AC004573
NID         g4204255
VERSION     AC004573.1  GI:4204255
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1  (bases 1 to 85095)
  AUTHORS  Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
           Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
           Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
           Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
           Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
           Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
           Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
           Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
           Rubin,G.M.
  TITLE    Sequencing of Drosophila chromosome 2L, region 21C5-21D1
  JOURNAL  Unpublished (1998)
REFERENCE  2  (bases 1 to 85095)
  AUTHORS  Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
           Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
           Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
           Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
           Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
           Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,P.,
           Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.P., Yee,A., Zhang,R.,
```

```
            Zderan I. I. and Kimmel B.E
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Jan 20, 1999 this sequence version replaced gi:4164118.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library location: 80-26.
FEATURES             Location/Qualifiers
     source          1..85095
                     /organism="Drosophila melanogaster"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="21C5-21D1"
                     /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
                     /clone="P1 DS07610 (Q275)"
BASE COUNT    22289 a   19671 c   20319 g   22816 t
ORIGIN


  Query Match             1.3%;  Score 24;  DB 37;  Length 85095.
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1312 catgaaactacgacaaagctgaagccg 1335
           ||||||||||||||||||||||||||||
Db    7600 CATGAACTACGACAAGCTGAGCCG 7577


RESULT 11
HS2P1_7
LOCUS       HS2P1_7                 1152-? bp    DNA             PRI       21-SEP-1998
DEFINITION  Human DNA sequence from clone 28C20 on chromosome 6q24. Contains a
            Zinc Finger gene ZNF131 pseudogene, ESTs an STS and GSSs, complete
            sequence.
ACCESSION   AL031119
NID         g3550036
VERSION     AL031119.1  GI:3550036
KEYWORDS    HTG; ZNF131.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 115835)
  AUTHORS   Smalley,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 8, 1998 this sequence version replaced gi:3355517.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 28C20. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre Chromosome 6
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            28C20 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..115835
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="28C20"
                     /map="q24"
                     /clone_lib="RPCI1"
     repeat_region   1..916
                     /note="L1 repeat: matches 3883..4802 of consensus"
     repeat_region   complement(1644..2446)
                     /note="L1MB5 repeat: matches 870...30 of consensus"
     repeat_region   complement(2455..2789)
                     /note="THE1C repeat: matches 356...1 of consensus"
     repeat_region   complement(2782..2956)
                     /note="L1 repeat: matches 5284...5111 of consensus"
     repeat_region   3865...4089
                     /note="MLT1F repeat: matches 317..541 of consensus"
     repeat_region   complement(8011...8367)
                     /note="THE1B repeat: matches 364...6 of consensus"
     repeat_region   complement(8641..9132)
                     /note="L1MA9 repeat: matches 1017..522 of consensus"
     repeat_region   complement(9154..10033)
                     /note="L1PA15 repeat: matches 904..1 of consensus"
     repeat_region   complement(9910..10804)
                     /note="L1 repeat: matches 5390...4471 of consensus"
     repeat_region   10805..10838
                     /note="17 copies 2 mer gt 94% conserved"
     gene            complement(11946..14372)
                     /gene="dJ28C20.1"
     CDS             complement(join(11946...13396,14220...14372))
                     /gene="dJ28C20.1"
                     /note="Zinc Finger gene ZNF131 pseudogene; match: proteins
                     P52739 Q90851 Q14526 P10074 Q15916 P24278 Q15260 Q15062
                     Q92010 Q08376 Q00403 Q13397 Q99592 P41183 P41182 Z13105
                     Q73453 Q15156 Q60821 Q43298 P70616 Q03172 P28160 P15822
                     Q15918 Q15044 Q05215 Q26452 Q24162 Q92981 Q35745 P36508
                     P10925 O60893 P18744 Q02313 Q02313 Q03924 Q14585 Q60493
                     Q14584 Q00403 Q08376 Q60821 Q73453 Q62981 Q62513 Q43693
                     Q13105 P51523 P52739; match: cDNA U09410; match: ESTs
                     F06005 R19845 R34147 F08171 AA492467 F12820 AA156200
                     AA482467 T75257"
                     /codon_start=1
                     /pseudo
                     /evidence=not_experimental
                     /db_xref="PID:e1321023"
     repeat_region   complement(13829..14060)
                     /note="LTR1 repeat: matches 224..1 of consensus"
     repeat_region   complement(14650..14787)
                     /note="LTR1 repeat: matches 674..524 of consensus"
     repeat_region   complement(15148..15332)
                     /note="LTR1 repeat: matches 185...1 of consensus"
     repeat_region   complement(15504..17889)
                     /note="L1 repeat: matches 4474..2060 of consensus"
     repeat_region   complement(17892..18190)
                     /note="AluSg repeat: matches 300..1 of consensus"
     repeat_region   complement(18192..19038)
                     /note="L1 repeat: matches 2071..1248 of consensus"
     repeat_region   complement(19045..19263)
                     /note="AluYb8 repeat: matches 308...90 of consensus"
     repeat_region   complement(19267..19864)
                     /note="L1 repeat: matches 1259...677 of consensus"
     repeat_region   complement(21144..21218)
                     /note="L1MA9 repeat: matches 529..455 of consensus"
     repeat_region   21810..21956
                     /note="MIR2 repeat: matches 1...146 of consensus"
     repeat_region   complement(23359..23408)
                     /note="MIR repeat: matches 262..214 of consensus"
     repeat_region   complement(24208...24350)
```

```
                     /note="MIR repeat: matches 156   15 of consensus"
     repeat_region   25536. .26822
                     /note="L1 repeat: matches 4106   5390 of consensus"
     repeat_region   26674. .27565
                     /note="L1PA2 repeat: matches 1. .893 of consensus"
     repeat_region   complement(27574. .27920)
                     /note="MER42C repeat: matches 1536. 1189 of consensus"
     repeat_region   28642. .28848
                     /note="L1ME1 repeat: matches 357   561 of consensus"
     repeat_region   complement(29041. .29382)
                     /note="AluSx repeat: matches 42  .1 of consensus"
     misc_feature    30490. .30933
                     /note="match: genomic DNA Z78623"
     repeat_region   30951. .31546
                     /note="MER5C repeat: matches 1515. .2124 of consensus"
     repeat_region   31452. .34748
                     /note="L1 repeat: matches 1431. .4813 of consensus"
     repeat_region   34749. .35046
                     /note="AluSq repeat: matches 1   300 of consensus"
     repeat_region   35048. .35628
                     /note="L1 repeat: matches 4796. .5388 of consensus"
     repeat_region   35484. .35901
                     /note="L1MA4 repeat: matches 1. .429 of consensus"
     repeat_region   36692. .36867
                     /note="MLT2_internal repeat: matches 953. .1131 of
                     consensus"
     repeat_region   36936. .37144
                     /note="MER5C repeat: matches 1. .217 of consensus"
     repeat_region   complement(37694. .38057)
                     /note="MLT1A2 repeat: matches 372   1 of consensus"
     repeat_region   38078. .38370
                     /note="MLT2_internal repeat: matches 1733. .2032 of
                     consensus"
     repeat_region   38363. .38592
                     /note="MLT2_internal repeat: matches 2221   2461 of
                     consensus"
     repeat_region   39064. .39677
                     /note="L1MB1 repeat: matches 288   916 of consensus"
     repeat_region   40912. .40943
                     /note="16 copies 2 mer aa 97% conserved"
     repeat_region   complement(44275. .44567)
                     /note="AluY repeat: matches 293. .1 of consensus"
     repeat_region   45637. .46082
                     /note="HAGGER1 repeat: matches 25. .474 of consensus"
     repeat_region   complement(46094. .46984)
                     /note="L1PA1 repeat: matches 904   1 of consensus"
     repeat_region   complement(46844. .50086)
                     /note="L1 repeat: matches 5390   2153 of consensus"
     repeat_region   50060. .52022
                     /note="HAGGER1 repeat: matches 430. .2418 of consensus"
     misc_feature    complement(52546. .53226)
                     /note="match: GSS B54551"
     repeat_region   52809. .52983
                     /note="MIR repeat: matches 36. .210 of consensus"
     repeat_region   complement(57666. .57826)
                     /note="MER5B repeat: matches 178   3 of consensus"
     repeat_region   58563. .58610
                     /note="24 copies 2 mer tt 81% conserved"
     repeat_region   complement(58612. .58912)
                     /note="AluSx repeat: matches 300. .1 of consensus"
     misc_feature    59207. .59752
                     /note="match: GSS B57434"
     repeat_region   complement(59910. .60566)
                     /note="L1PA8 repeat: matches 910   256 of consensus"
     misc_feature    60824. .61454
                     /note="match: GSS B99968"
     repeat_region   63492. .64525
                     /note="17 copies 2 mer tg 97% conserved"
     misc_feature    complement(64026. .64385)
                     /note="match: GSS AQ089270"
     repeat_region   complement(65072. .65381)
                     /note="AluSp repeat: matches 303   1 of consensus"
     repeat_region   complement(65755. .65825)
                     /note="MIR repeat: matches 146   76 of consensus"
     misc_feature    complement(66137. .66432)
                     /note="match: GSS AQ066992"
     repeat_region   complement(68310. .68882)
                     /note="MLT2D repeat: matches 553. .1 of consensus"
     repeat_region   69187. .69291
                     /note="MIR repeat: matches 20. .141 of consensus"
     misc_feature    72197. .72557
                     /note="match: GSS B90081"
     repeat_region   72893. .73320
                     /note="MER28 repeat: matches 1   434 of consensus"
     repeat_region   75287. .75648
                     /note="THE1C repeat: matches 1   371 of consensus"
     repeat_region   complement(76395. .76458)
                     /note="MIR2 repeat: matches 134   70 of consensus"
     repeat_region   complement(76658. .76719)
                     /note="MLT2G repeat: matches 388. .324 of consensus"
     repeat_region   complement(76658. .76706)
                     /note="MLT2D repeat: matches 553. .505 of consensus"

  Query Match              1.2%;  Score 22;  DB 9;  Length 115835;
  Best Local Similarity   100.0%;  Pred. No. 0.4;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1525 ctgctctgaccttccagagccc 1546
           ||||||||||||||||||||||
Db   88123 CTGCTCTGACCTTCCAGAGCCC 88102


RESULT  12
MMNETRN
LOCUS       MMNETRN      2658 bp    mRNA            PRD       13-MAR-1995
DEFINITION  M.musculus net mRNA.
ACCESSION   Z32815
NID         g479112
VERSION     Z32815.1  GI:479112
KEYWORDS    Net; ras gene.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2658)
  AUTHORS   Giovane,A., Pintzas,A., Maira,S.M., Sobieszczuk,P. and Wasylyk,B.
  TITLE     Net, a negative factor switch to positive by Ras
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2658)
  AUTHORS   Giovane,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1994) Antoine Giovane
            CNRS-LGME,INSERM-U.184,Institut de Chimie, Biologique, 11 rue
            Humann, Strasbourg, 67085 Strasb. Cedex, France
REFERENCE   3  (bases 1 to 2658)
  AUTHORS   Giovane,A., Pintzas,A., Maira,S.M., Sobieszczuk,P. and Wasylyk,B.
  TITLE     Net, a new ets transcription factor that is activated by Ras
  JOURNAL   Genes Dev. 8 (13), 1502-1513 (1994)
  MEDLINE   95047310
FEATURES             Location/Qualifiers
     source          1. .2658
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /dev_stage="10 days"
                     /tissue_type="embryo"
                     /clone_lib="lambda zap2"
     CDS             295   1524
                     /codon_start=1
                     /product="Net"
                     /protein_id="CAA83676.1"
                     /db_xref="PID:g479113"
                     /db_xref="GI:479113"
                     /db_xref="MGD:MGI:101762"
                     /db_xref="SWISS-PROT:P41971"
                     /translation="MESAITLWQFLLHLLLDQKHEHLICWTSNDGEFKLLKAEEVAKL
                     WGLRKNKTNMNYDKLSRALRYYYDKNIIKKVIGQKFVYKFVSFPDILKMDPHAVEISP
```

```
                     ESLLLQDGDCKVSPEGREVHRHGLSSLKSASRNEYLHSGLYSSFTINSLENAPEAFKA
                     IKIEKLEEPCDDSPPVEEVRTVIRFVTNKTDKHITRPVMSLPSTSETAAAAASAFLAS
                     SVSAKISSLMLPNAASVSSVSPSSSSRSPSLSPDSPLPSEHPSLFLFAACHESDSLEPL
                     NLSSGSKIKSPSLPPKGKKPKGLEISAPQLLLSGTDIGSIALNSPALPSGSLTPAFFT
                     AQTPSGLFLASSPLLPSIHFWSSLSPVAPLSPARLQGPNTLFQFPTLLNGHMPVPLPS
                     LDRAPSPVLLSPSSQKS"
BASE COUNT      682 a    714 c    600 g    662 t
ORIGIN


  Query Match             1.2%;  Score 22;  DB 12;  Length 2659;
  Best Local Similarity   100.0%;  Pred. No. 0.39;
  Matches   22;  Conservatives    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1312 catgaactacgacaagctgagc 1333
           |||||||||||||||||||||
Db    453 CATGAACTACGACAAGCTGAGC 474


RESULT  13
MUSTRANON1
LOCUS       MUSTRANONC    2211 bp    mRNA            ROD       19-JUL-1994
DEFINITION  Mus musculus ERP mRNA, complete cds.
ACCESSION   L19953
NID         g436184
VERSION     L19953.1  GI:436184
KEYWORDS    transcription factor.
SOURCE      Mus musculus (strain BALB/c, sub_species domesticus) (library:
            lambda-gt11 of Clontech) male lung cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2211)
  AUTHORS   Lopez,M., Oettgen,P., Akbarali,Y., Dendorfer,U. and Libermann,T.A.
  TITLE     ERP, a new member of the ets transcription factor/oncoprotein
            family: cloning, characterization, and differential expression
            during B-lymphocyte development
  JOURNAL   Mol. Cell. Biol. 14 (5), 3292-3309 (1994)
  MEDLINE   94217726
FEATURES             Location/Qualifiers
     source          1. .2211
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /sub_species="domesticus"
                     /db_xref="taxon:10090"
                     /sex="male"
                     /tissue_type="lung"
                     /tissue_lib="lambda-gt11 of Clontech"
     gene            302. .2199
                     /gene="ERP"
     CDS             302. .1531
                     /gene="ERP"
                     /codon_start=1
                     /protein_id="AAA19872.1"
                     /db_xref="PID:g436185"
                     /db_xref="GI:436185"
                     /translation="MESAITLWQFLLHLLLDQKHEHLICWISNDGEFKLLKAEEVAKL
                     WGLRKNKTNMNYDKLSRALRYYYDKNIIKKVIGQKFVYKFVSFPDILKMDPHAVEISR
                     ESLLLQDGDCKVSPEGREVHRHGLSSLKSASRNEYLHSGLYSSFTINSLQNAPEAFKA
                     IKTEKLEEPCDDSPPVEEVRTVIRFVTNKTDKHITRPVMSLPSTSETAAAAASAFLAS
                     SVSAKISSLMLPNAASVSSASPSSSSRSPSLSPDSPLPSEHRSLFLEAACHESDSLEPL
                     NLSSGSKTKSPSLPPKGKKPKGLEISAPQLLLSGTDIGSIALNSPALPSGSLTPTFFT
                     AQTPSGLFLASSPLLPSIHFWSSLSPVAPLSPARLQGPNTLFQFPTLLNGHMPVPLPS
                     LDRAPSPVLLSPSSQKS"
     misc_feature    320. .553
                     /gene="ERP"
                     /note="ETS domain"
     polyA_signal    2145. .2199
                     /gene="ERP"
BASE COUNT      560 a    613 c    504 g    534 t
ORIGIN


  Query Match             1.2%;  Score 22;  DB 12;  Length 2211;
  Best Local Similarity   100.0%;  Pred. No. 0.39;
  Matches   22;  Conservatives    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1312 catgaactacgacaagctgagc 1333
           |||||||||||||||||||||
Db    460 CATGAACTACGACAAGCTGAGC 481


RESULT  14
S82864
LOCUS       S82864       1526 bp    mRNA            ROD       10-FEB-1997
DEFINITION  Elk-3=Ets transcription factor [mice, 16-day embryos, mRNA, 1526
            nt].
ACCESSION   S82864
NID         g1836129
VERSION     S82864.1  GI:1836129
KEYWORDS    .
SOURCE      Mus sp. 16-day embryos.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1526)
  AUTHORS   Nozaki,M., Onishi,Y., Kanno,N., Ono,Y. and Fujimura,Y.
  TITLE     Molecular cloning of Elk-3, a new member of the Ets family
            expressed during mouse embryogenesis and analysis of its
            transcriptional repression activity
  JOURNAL   DNA Cell Biol. 15 (10), 855-862 (1996)
  MEDLINE   97047916
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 179297] from the original journal article.
            This sequence comes from Fig. 1.
FEATURES             Location/Qualifiers
     source          1. .1526
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
     gene            94. .1323
                     /gene="Elk-3"
     CDS             94. .1323
                     /gene="Elk-3"
                     /note="Ets transcription factor, This sequence comes from
                     Fig. 1"
                     /codon_start=1
                     /product="Elk-3"
                     /protein_id="AAB46893.1"
                     /db_xref="PID:g1836130"
                     /db_xref="GI:1836130"
                     /translation="MESAITLWQFLLHLLLDQKHEHLICWISNDGEFKLLKAEEVAKL
                     WGLRKNKTNMNYDKLSRALRYYYDKNIIKKVIGQKFVYKFVSFPDILKMDPHAVEISR
                     ESLLLQDGDCKVSPEGREVHRHGLSSLKSASRNEYLHSGLYSSFTINSLQNAPEAFKA
                     IKTEKLEEPCDDSPPVEEVRTVIRFVTNKTDKHITRPVVSLPSTSETAAAAASAFLAS
                     SVSAKISSLMLPNAASVSSASPSSSSRSPSLSPDSPLPSEHRSLFLEAACHESDSLEPL
                     NLSSGSKTKSPSLPPKGKKPKGLEISAPQLLLSGTDIGSIALNSPALPSGSLTPTFFT
                     AQTPSGLFLASSPLLPSIHFWSSLSPVAPISPARLQGPNTLFQFPTLLNGHMPVPLPS
                     LDRAPSPVLLSPSSQKS"
BASE COUNT      353 a    487 c    350 g    336 t
ORIGIN


  Query Match             1.2%;  Score 22;  DB 12;  Length 1526;
  Best Local Similarity   100.0%;  Pred. No. 0.39;
  Matches   22;  Conservatives    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1312 catgaactacgacaagctgagc 1333
           |||||||||||||||||||||
Db    252 CATGAACTACGACAAGCTGAGC 273


RESULT  15
PFMAL3P4/c
LOCUS       PFMAL3P4   113899 bp    DNA             INV       29-APR-1999
DEFINITION  Plasmodium falciparum MAL3P4, complete sequence.
ACCESSION   AL008970
```

```
NID         g4409924
VERSION     AL008970.2  GI:4409924
KEYWORDS    HTG.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
REFERENCE   1  (bases 1 to 113899)
  AUTHORS   Hamlin,N., Bowman,S., Churcher,C., Lawson,D., Quail,M. and
            Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (15 APR 1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Mar 24, 1999 this sequence version replaced gi:3763998.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum/MPOSTANI. This
            sequence is not the entire insert of clone MAL3P4. It may be
            shorter because we only sequence overlapping sections once, or
            longer because we arrange for a small overlap between neighbouring
            submissions.
FEATURES             Location/Qualifiers
     source          1. .113899
                     /organism "Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="3"
                     /clone="MAL3P4"
     gene            2364. .6248
                     /gene="MAL3P4.26"
     CDS             2364. .6248
                     /gene="MAL3P4.26"
                     /note "predicted using hexExon; MAL3P4.26 (PFC0435w),
                     Hypothetical protein, len: 1295 aa"
                     /codon_start=1
                     /protein_id "CAA15616.1"
                     /db_xref="PID:e1332567"
                     /db_xref "PID:g3764023"
                     /db_xref "GI:3764023"
                     /db_xref "SPTREMBL:O77361"
                     /translation="MKVGIIFFCLFFFVVLGACNNVKEKIFKNIKKKTKFIILNEPIV
                     DLSFSENLFHTLLFDLDVDKNLYTLDESLLNLENLNYSSIFRLLVDTYKNIKENEDDN
                     KNIRYIFLCTSFSRIHPLNFEYFLRKLNKYIYNGNIYEKGNVDIRGILEEYNKEIEEK
                     KLEKQKLNKIKDKNNNNNNNNNSKFSKDGDNEDFNNKNDLYNPSDKLYNNNDDIDVHE
                     LLEEIIIKEKRFFLNDDDDNDSNDKYILKTDEVNKYKGFFIGYGFNDDIPSVIHHYNF
                     DKNFLFFSLNSGIILDITLLKNIYEVSNILLSNNEKDQSIHIDYIYEVTKYIKENLRV
                     RLIHSENVCLNEEQNIHLLDNDPNNFEIYKYYQVLNLFKDYNKNTEEKQYEKIGHENV
                     RHEETSSEGNENLNRNTKHNNDNNNDNNNYSEDAIAELLLSYFNVFYPISTCMCYSIR
                     SKHESLMDYDKYHMINLENDIKLKHYIKETEEIHFNSIEEYKMKLNKINYKYDILLEE
                     HENLVTHKNILIGIKTSINTEEERIPHIKNTYDNKENTQIIFNTFNYDNKLKEKNTFG
                     FYNNSLLQNALENDNIDLDIIYMSDKESQKYDNLYFNSKVTSKEGLCEKLKHMIYYYY
                     EEYVMKNSEKKYFEIADDDIFVNVKNLIDVTNLTLNTCSHSKKYMYDKYIKSYDFVKE
                     NEALFLQNFDKKILFLYSYLKDTFAKTIQTLKKYDYVPKYCQGGILSKKHKNNDSDDD
                     HDHHVGNKQNNDSTNHQDIEKNQVNVINNNNNNNNNKAKSIPIYLGRRYSYNTFSTNS
                     NEYFYDYLFGGAGILINDETAKRIYECKECTCPSTNSSMDDMIFGKWAKELGILAINF
                     EGYFQNSPLDYNKKYINFLVPIIYHRLNKNRTTKESRDMYFNYLVNYNRNDKFQNKDI
                     YVDYLLDRNHKNMIPNVFHYFFYVNMYDEKNKVVTKIEHNADMNSKKNKSKNPQKLNNT
                     QCDKNVNIDDNVNIDDENVKGDENVKGDENVKGDEYMKGDENVKGDENVKDDENVKDDE
                     NIKGDDNNYNVDNMENIDDIINMVESVDDDVMERNKKGTGKEKKDDKNHNNKEKATDV
                     KKSSVGNNIQKNFDIIKYVIKMNEKIYNRMQESGKYKQLFDINKFFKKEIEGHPYFQ
                     KIKKKNEEAKKKLKMNQLKKQKDYINNYFHTSNMQGNFNQQKMGNYQNQENEENDFF
                     DQDFEIEEDAINPMDYEEYMENLSNFEDDGEPYDEYDDYDDFVNTINADKLKINDQNK
                     HLYEQIKDIAQPPVNFQNDQNSNTFDFDTDEL"
     gene            complement(7161. .13970)
                     /gene="MAL3P4.25"
     CDS             complement(7161. .13970)
                     /gene="MAL3P4.25"
                     /note="predicted using hexExon; MAL3P4.25 (PFC0440c),
                     Helicase, len: 2270 aa; Similarity to helicases. C.elegans
                     helicase (WP:F52B5.3) BLAST Score: 290, sum P(4) =
                     5.9e-41; 44% in 199 aa overlap."
                     /codon_start=1
                     /protein_id "CAA15615.1"
                     /db_xref "PID:e1332566"
                     /db_xref="PID:g3764022"
                     /db_xref="GI:3764022"
                     /db_xref="SPTREMBL:O77360"
                     /translation="MNNTFKINKKRRTVYESYNIINIANKKRYMCNDNKMDNDNNIIN
                     EKKYFNNSSSIKNTGNNYKNDNIDKSYDMCERRSMMDRDENIYNNNYKNKNRNNYSNK
                     NNYSNNINHMDNRRNTYKGDLDGKKQYLSNKHYESKNDNENSYNLKKYNFNIFFIKYN
                     NNNNNNNNNNNNDPQNIDNNIYKFYDDKNEKKSNKTHSYGNNFNNEFNTENDPKQNYM
                     SNESHNMTQDGHPIYNHNNNNNDGRIKNRKDISLSSYGYNKISNDNLSKENTSYINQY
                     NNSGGANRHVINAQHNNNQHNNNQHNNNFNYHNYKNAKPYQDKYKNGDSFKFSAYNAE
                     QHDNMDKNYYQDIHKNNMKHSTKGDTAISFNGQKVFGSFKNKQMKEKRKFFKNKERGS
                     SINRPGMHNNNNNYNNNNNNYYNNYNNHFCDDAERRNNKNENWCEINSNELYEKDGD
                     MPIYNANENCNLKKMIYKNDKKEKEKIKILLTEDNVNNVQSNIEKLSIYKSRNEIIEM
                     IEKNDVTFINGETGSGKSTCVPKFLLEENIRENKKINIIVTEPRRIACIALSKILSEL
                     TNEKLGQKIGYRISGESLYDSEKTVIIYIIDGYLFKLFIHHKNMYKKFTHVIIDFIHD
                     RSILLDIVLLFIKLYLHNKQKDEQMFKLIIMSATMQSNLFYSYFEHPNIKMGSIFIGT
                     KIFSIDTFYIEDIINYTFYGSRNIHDFKINDFKINDFKINDFKINDFKINDNMFSDD
                     DKKSCDDKKSCDGNNNFVKESVIDFILRRKNCNKINLSKNSEMLLYKIKSEYDKNIHM
                     FNNNNNKYCNDKNKDLDVDEIIPANVFSNISNLCLELVYNLCLKGDSVIIFLSGMQDI
                     TDMYHQLSMIINNNMDNPSNSNININNGNDANDANDVNNMHNMHNMHNMHNIANQLTY
                     HRTDIKIHIHMLHSCLYDNTIHKLKHNDTDINIFLSSNIAESSITIPNVRLVIDFCIQ
                     KNIEYNDKKAHILVKKWINKSSMEQRKGRCGRTCHGICIRMISKNFINLIRDHKISE
                     IYTHSLHLLYLYILKSMSVLNGLINKRNEVVHCDNTNVYNVKEVKDQKNGICKNEISH
                     SHAENKKLTIYDVLSMIIEKPSKEKIKSIRYFLEKVKAVIKIKDKLVISIGQIMIRF
                     NLSINLCRLLLYGVLLDVTFDTIILIGILNTNDIFPNINLYSSKNIYSYAVSLEICSK
                     QKAYFDGNTYSEPIMLRNVFLEWLCVYLLYVQGLKKENKFHRKELKTYYMNTCSIMNK
                     RNHVNSKKLLCVINSVDNLCKKMLKMLNKNSNAYKSCVYLLYLLRGGSDINYNMGNNS
                     NNNNDNSGEIINADGITPNNIHMNEINGMNVFNVVTKYCLFDYSNQNIYLKFLFSLSF
                     TPLFIHGSPNLPLRDLNEGKKKSKKLMNVLNFMIDNKLDVKNCIYFSGVYIPDINILK
                     RAICIMCPYLSFDIYVSKNIYYVYFHNNNYHNNYHNNILKGDDFSYYNKTICDKGDAD
                     MSVNSYLNLTQEGHDTDGITSFGVSPLMNNNNMYGNNYIHNNVYYPCNGINNYYNNNV
                     VNSQNNYMMKMIMSNYYSESFLNLTSGKNYPDIIEKIPNKLLYKYNNNVEKSIFHNFY
                     NNILNLIQPVNAIVPVYSNKYDEINNASICTNIINMFSNGKWIFTMPIFNPNKRKCEV
                     SEYFNHKYELKKPKHLFLVKWMFLNTDNYKIKKNKQKEHVMDINNALVNNNSLGKNVS
                     KNDIHRDNINDYSLNDQEDQQYNKGYYYIDDYDDYDDYDDDNNDDNNDNNNNNNDDNN
                     NNNDDDYLSNVKVESDNNNTHQNNPDGYEGNDSSSMCSDNNKKSKQKKMKCNLNFPSV
                     LGFLSMCPFNFDTKRNIYDNQTADIYAVCSSIDYSYINDTYTWVNYVIVLPNKYFLTF
                     FLSSIPYHDNVIIQTPTNLYTSDILSIKIFDSKEIILKNMKKMKHKNKTNTTNYYDTN
                     STINNPNITKHDLLFVNYLPYVMSKLLLSLTLSFNPQEIEQNTQPLKLSEKSKRKSVI
                     INKDGPNNDNFVDANDKNVNGNDKNVNDNDKNVNSQDNQGNQKDHFDDEDEEYFSDED
                     IMSDIEENNILKENLYLKQIVINMIEENNMDYINYNKYSMDNMDIDEHFNLQYLYEDE
                     HFINLNEKWYNNKGEILMNTYDYLYNSLKNENDYNQYNNNNNNNNNNNNSSYMVKNI
                     KEKALNSLEAVKSSYLNNNFSVQVQNTMNAYERINSEDFLFFQPINISPIKEANYKLR
                     SMYYNEVAPTLDKFRRTNRQV"
     gene            18825. .19157
                     /gene="MAL3P4.24"
     CDS             18825. .19157
                     /gene="MAL3P4.24"
                     /note="predicted using hexExon; MAL3P4.24 (PFC0445w),
                     Hypothetical protein, len: 111 aa"
                     /codon_start=1
                     /protein_id="CAA15613.1"
                     /db_xref="PID:e1332564"
                     /db_xref="PID:g3764020"
                     /db_xref="GI:3764020"
                     /db_xref="SPTREMBL:O77358"
                     /translation="MLHGISTISEKINVYSLYEKKTNIFQSLEKKGIETIFGDGFKIQ
                     CYDTLTGIKIFIVHKDDLNIEMNTYIKRVYELYSDIILKNPFYDIDMPIRSAVFNFQI
                     EKLFSNIS"
     gene            19545. .19874
                     /gene="MAL3P4.23"
     CDS             19545. .19874
                     /gene="MAL3P4.23"
                     /note="predicted using hexExon; MAL3P4.24 (PFC0450w),
                     Hypothetical protein, len: 110 aa"
                     /codon_start=1
                     /protein_id="CAA15614.1"
                     /db_xref="PID:e1332565"
                     /db_xref="PID:g3764021"
                     /db_xref="GI:3764021"
                     /db_xref="SPTREMBL:O77359"
                     /translation="MSMFLNILILIDAASVAFLLITFLMINLNEESLELSQAHRENGK
                     KALVVAIILYVIFLVLLFIYKAYKNKRKLYTNFFMKKRNAPKYVQLASIYLSASDEYE
                     QYELNKI"
```

```
     gene            20969. .22229
                     /gene="MAL3P4.22"
     CDS             20969. .23328
                     /gene="MAL3P4.22"
                     /note="predicted using hexExon; MAL3P4.22 (PFC0455w),
                     Hypothetical protein, len: 787 aa"
                     /codon_start=1
                     /protein_id="CAA15612.1"
                     /db_xref="PID:e1332563"
                     /db_xref="PID:g3764019"
                     /db_xref="GI:3764019"
                     /db_xref="SPTREMBL:O77357"
                     /translation="MKLLKLDYYNGEIYEKYKINNILSLLVDNNKFQNESVDFHEISN
                     KKNNYSIMAILGSQIKNLNDDIYKDIDTNDIILALSKLKKTDEEILINESQNKNYILE
                     LAEQIYVFYFQPGIIKNDHNFHLNNSKKNFIQINIELMNEINPEIKNYEHKIHNLPAY
                     SFNAQTKNDASINDLLMYYHEDVDNTDDSLLNENNFNYNMKHKKYNDQNKIFLNDHAH
                     NNNHNNNNNDFFDRFNENNFLMDEENVLKAAHEDIPMNNDYVNYIHMKRDSLHNNNNN
                     NNNNNYNNNNYNNYKNNIFSDHHDLFHYNKEGTNSTNCNSVENLNYDINIYKNNKKNI
                     QHINNNVNSNENIPFVSNFNGNIGNDGFNSYFNLSQLNEEDKHLLYNDTIKMDYLLNF
                     KIEYEKLDEDQSESNNKSMMVMNTHRMMKNFDNMNKINSNVVDTYEENMNYKYDKNKS
                     YVLNDEETYEHMSNKNMNNVMKNTGANNGANNGVEENCKYHIDMDNVHNAVNSVNLQD
                     NIYKNNNSSNTLNSVLENIDKAEQQIVKNHLKNSLAHNEYVAKRMSSLLLNPQNVIDK
                     EPYNNNIFKAYEKNGIMNRDEQKYFNDEISLNKYSSTYGQKKNFTTDTNNSSSFILST
                     IHKPNMVDESSYNIMNNYENKKFRIQSRNMNTRENYLHNQNSNHHSTERSHNINNNMI
                     NHNNHISNTIPLNFDNMSNYSRNAAKSVESNMNIYTNKYSRALDINDLNSNNTSRNYV
                     NSSNSLKSDLIMMRDNYAKMLERSKSLTSHLYRAPIMHNKYLKANM"
     gene            24170. .25001
                     /gene="MAL3P4.21"
     CDS             join(24170. .24420,24635. .25001)
                     /gene="MAL3P4.21"
                     /note="predicted using hexExon; MAL3P4.21 (PFC0460w),
                     Hypothetical protein, len: 206 aa"
                     /codon_start=1
                     /protein_id="CAA15611.1"
                     /db_xref="PID:e1332562"
                     /db_xref="PID:g3764018"
                     /db_xref="GI:3764018"
                     /db_xref="SPTREMBL:O77356"
                     /translation="MYTYIITYLNIYFIKIRMMMNPTLGKPEEDLAVYDYTDISMNVCL
                     KKYEYIPDPWINDSPYIYSNENMQNEIEALNELMKSKVSLVCAFFNPAFSYLFMTIGF
                     IGMFVFYKMKNDLSMGFIFACIFLVSFSHYIMCIICSYIITKDVIKSVEIRIKSLNIK
                     YKKEKIVPKLINLSSFGLKYLKLFTLSKGQAYPTLRVIYKYCKII"
     gene            complement(26011. .28044)
                     /gene="MAL3P4.20"
     CDS             complement(join(26011. .27209,27372. .27599,27795. .28044))
                     /gene="MAL3P4.20"
                     /note="predicted using hexExon; MAL3P4.20 (PFC0465c),
                     Hypothetical protein, len: 559 aa"
                     /codon_start=1
                     /protein_id="CAA15610.1"
                     /db_xref="PID:e1332561"
```

Query Match 1.2%; Score 22; DB 36; Length 113899;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1242 aaaaaaaaaaaggggcatcttca 1263
           |||||||||||||||||||||||
Db   49154 AACAAAGAAAAGGGCATCTTCA 90282
```

Search completed: November 20, 1999, 15:27:56
Job time: 66s sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 1999, 15:16:47 ; Search time 505.23 Seconds
(without alignments)

11922.305 Million cell updates/sec

Title: US-09-126-945-1
Perfect score: 1894
Sequence: 1 gtctgacttcctcccagcac ataaagatactagagaactg 1894

Scoring table: OLIGODX_NUC

Searched 679419 seqs, 1538154680 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1894	100.0	1894	11	AF071538	AF071538 Homo sapi
2	38	2.0	1704	12	AB019436	AB019436 Mus muscu
3	30	1.6	848	4	AF075706	AF075706 Gallus ga
4	29	1.5	2508	4	AF075708	AF075708 Gallus ga
c 5	28	1.5	60056	35	AC007818	AC007818 Drosophil
6	28	1.5	362	36	DRODETS4A	M88474 Drosophila
7	25	1.3	1403	10	HSDNAFEV3	Y08979 H.sapiens F
8	25	1.3	1901	10	HSRNAFEV	Y08976 H.sapiens m
9	24	1.3	480	36	DRODETS6A	M88475 Drosophila
c 10	24	1.3	85095	37	AC004573	AC004573 Drosophil

```
c  11      22    1.2  115935   9  HS28220                  AL031119 Human DNA
   12      22    1.2    2658  12  MMNETRN                   Z32815 M.musculus
   13      22    1.2    2211  12  MUSTRANONC                L19953 Mus musculu
   14      22    1.2    1529  12  S82864 Elk-3-Ets t        S82864 Elk-3-Ets t
c  15      22    1.2  114896  36  PFMAL3P4                  AL008970 Plasmodiu
   16      21    1.1  133453  34  HS989H11                  Z83851 Homo sapien
   17      20    1.1  129569   1  D90912                    D90912 Synechocyst
   18      20    1.1     473   5  E06024                    E06024 cDNA encodi
   19      20    1.1    2073   5  E06025                    E06025 cDNA encodi
c  20      20    1.1   36742   9  D87002                    D87002 Human (lamb
   21      20    1.1    2064   9  HUMELAF                   D12765 Human mRNA
c  22      20    1.1  152141  10  HSU07000                  U07000 Human break
   23      20    1.1    2667  10  HSU15655                  U15655 Human ets d
   24      20    1.1    2443  10  HSU18018                  U18018 Human E1A e
   25      20    1.1   35687  11  AC000051                  AC000051 Homo sapi
   26      20    1.1   34539  11  AC002051                  AC002051 Homo sapi
c  27      20    1.1  151111  11  AC002308                  AC002308 Homo sapi
c  28      20    1.1   36921  11  AC004185                  AC004185 Homo sapi
c  29      20    1.1  112220  11  AC005373                  AC005373 Homo sapi
c  30      20    1.1   37139  11  AC006047                  AC006047 Homo sapi
c  31      20    1.1  111084  11  AC006486                  AC006486 Homo sapi
   32      20    1.1    2642  11  HSETSF5                   AF095890 Homo sapi
   33      20    1.1    1946  12  MUSETSRPA                 L10427 Mus musculu
c  34      20    1.1  126093  34  AC005283                  AC005283 Homo sapi
c  35      20    1.1  230822  34  D84394                    D84394 Homo sapien
c  36      20    1.1  131436  34  HS87L1C4                  AL078476 Homo sapi
   37      20    1.1  269359  34  HSDJ94E24                 AL050317 Homo sapi
c  38      20    1.1  160079  35  AC007664                  AC007664 Homo sapi
c  39      20    1.1  102453  35  AC007708                  AC007708 Homo sapi
   40      20    1.1   58839  35  AC007821                  AC007821 Drosophil
   41      20    1.1    1832  37  AF106964                  AF106964 Strongylo
   42      20    1.1   47881  42  AC002054                  AC002054 Homo sapi
c  43      20    1.1  155413  42  AC006453                  AC006453 Homo sapi
   44      19    1.0   35516   1  MTCY373                   Z73419 Mycobacteri
   45      19    1.0   40790   1  MTCY493                   Z95844 Mycobacteri
```

ALIGNMENTS

```
RESULT  1
AF071538
LOCUS       AF071538      1894 bp    mRNA            PRI       13-DEC-1998
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
NID         g4007417
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Akbarali,Y., Thamrongsak,U., Dube,A.,
            Boltax,J., Brown,L., Kunsch,C. and Libermann,T.A.
  TITLE     Isolation and characterization of a novel prostate
            epithelium-specific Ets transcription factor, PDEF
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium-specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="PID:g4007418"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSPTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFPQPSPLGGDVLHAHLDIWKSAAWMKEP
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWINKEK
                     GIFKIEDSAQVAPIWGIPKNPPAMNYDKLSRSIRQYYKKGIIRKPDISQRIVYQFVHP
                     I"
BASE COUNT      368 a    653 c    571 g    302 t
ORIGIN
```

```
  Query Match           100.0%;  Score 1894;  DB 11;  Length 1894;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1894;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtctgacttcctcccagcacattcctgcactctgccgtgtccacactgcccccacaaacc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTCTGACTTCCTCCCAGCACATTCCTGCACTCTGCCGTGTCCACACTGCCCCACAGACCC 60

Qy      61 agtcctccaagcctgctgccagctccctgcaagcccctcaggttgggccttgccacggtg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGTCCTCCAAGCCTGCTGCCAGCTCCCTGCAAGCCCCTCAGGTTGGGCCTTGCCACGGTG 120

Qy     121 ccagccaggcagccctgggctgggggtagggagctccctacaggcacgcagccctgaagcc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCAGCAGGCAGCCCTGGGCTGGGGGTAGGGAGCTCCCTACAGGCACGCAGCCCTGAGACC 180

Qy     181 tcagagggccacccccttgagggtggccaggcccccagtggccaacctgagtgctgcctct 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCAGAGGGCCACCCCTTGAGGGTGGCCAGGCCCCCAGTGGCCAACCTGAGTGCTGCCTCT 240

Qy     241 gccaccagccctgctggccccctggttccgctggcccccccagatgcctggctgagacacgc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCCACCAGCCCTGCTGGCCCCCTGGTTCCGCTGGCCCCCCAGATGCCTGGCTGAGACACGC 300

Qy     301 cagtggcctcagctgcccacacctcttcccggcccctgaagttggcactgcagcagacag 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGTGGCCTCAGCTGCCCACACCTCTTCCCGGCCCCTGAAGTTGGCACTGCAGCAGACAG 360

Qy     361 ctccctgggcaccaggcagctaacagacacagccgccagccccaaacagcagcggcatggg 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTCCCTGGGCACCAGGCAGCTAACAGACACAGCCGCCAGCCCAAACAGCAGCGGCATGGG 420

Qy     421 cagcgccagccccgggtctgagcagcgtatccccagccagcctcctgctgccccccgacac 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAGCGCCAGCCCCGGGTCTGAGCAGCGTATCCCCAGCCAGCCTCCTGCTGCCCCCCGACAC 480

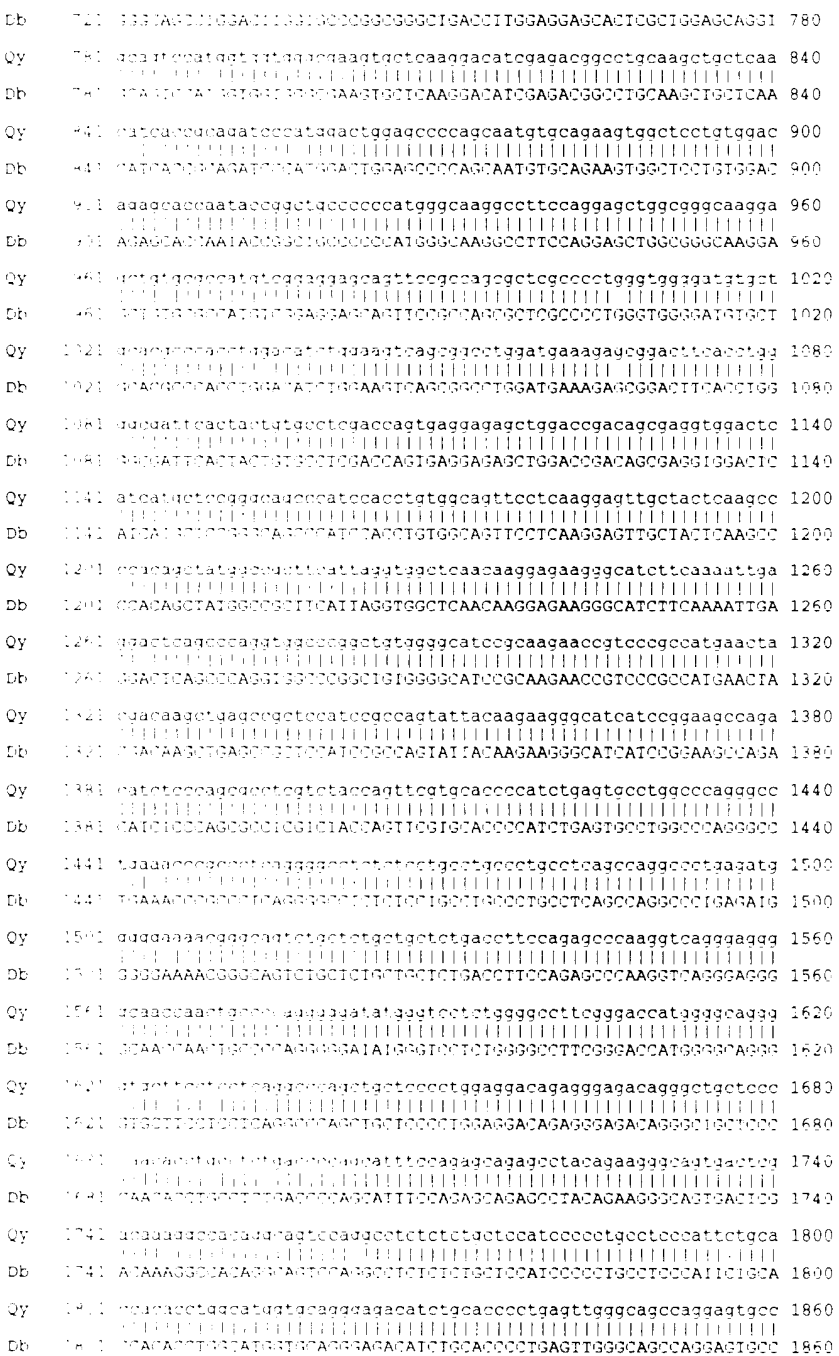
Qy     481 ggtgtcgcggacaggcttggagaaggcggcagccgggggcagtgggtctcgagagacccgga 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GGTGTCGCGGACAGGCTTGGAGAAGGCGGCAGCGGGGGCAGTGGGTCTCGAGAGACGGGA 540

Qy     541 ctggagtcccagtccacccgccacgcccgagcagggcctgtccgccttctacctctcctta 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CTGGAGTCCCAGTCCACCCGCCACGCCCGAGCAGGGCCTGTCCGCCTTCTACCTCTCCTA 600

Qy     601 ctttgacatgctgtacctgaggacagcagctgtgggcagccgagggcctgaggccgacag 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CTTTGACATGCTGTACCTGAGGACAGCAGCTGTGGGCAGCCGAGGGCCTGAGGCCGACAG 660

Qy     661 tcgggaggagccacctgaggagcctgagcagtgccccgtcattgacagccaaatccccagc 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TCGGGAGGAGCCACCTGAGGAGCCTGAGCAGTGCCCCGTCATTGACAGCCAAATCCCCAGC 720

Qy     721 gggcagcctggacttggtgcccggcgggctgacctgggaggagcactcgctgaaacaggt 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```



```
Qy   1861 cccgggaatggataataaagatactagagaaactg 1894
          ||||||||||||||||||||||| |||||||||
Db   1861 CCCGGGAATGGATAATAAAGATACTAGAGAACTG 1894


RESULT   2
AB019436
LOCUS       AB019436     1704 bp    mRNA            ROD       08-MAY-1999
DEFINITION  Mus musculus mRNA for Prostate specific transcription factor ets,
            complete cds.
ACCESSION   AB019436
NID         g4760536
VERSION     AB019436.1  GI:4760536
KEYWORDS    Prostate specific transcription factor ets; mPse.
SOURCE      Mus musculus (strain:C57BL/6) Adult male Colon Epithelial cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Prostate specific ets
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1. .1704
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /cell_type="Epithelial"
                     /dev_stage="Adult"
                     /sex="male"
                     /tissue_type="Colon"
     gene            362. .1339
                     /gene="mPse"
     CDS             362. .1339
                     /gene="mPse"
                     /codon_start=1
                     /product="Prostate specific transcription factor ets"
                     /protein_id="BAA77329.1"
                     /db_xref="PID:d1041095"
                     /db_xref="PID:g4760537"
                     /db_xref="GI:4760537"
                     /translation="MGSASPGLSNVSPGCLLLFPDVAPRTGTEKAASGAMGPEKQEWS
                     PSPPATPEQGLSAFYLSYFNMYPDDSSWVAKVPEARAGEDHPEEPEQCPVIDSQASGS
                     TLDEHSLEQVQSMVVGEVLKDIETACKLLNITADPGDWSPGNVQKWLLWTEHQYRLPP
                     AGKAFQELGGKELCAMSEEQFRQRAPLGGDVLHAHLDIWKSAAWMKEKISPGILHYCA
                     STSEEGWTDGEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWINKEKGIFKIEDSAQ
                     VARLWGVRKNPPAMNYDKISRSIPQYYKKGIIRKPDISQRLVYQFVHPV"
     polyA_site      1704
                     /note="10 a nucleotides"
BASE COUNT      369 a    527 c    491 g    317 t
ORIGIN


  Query Match             2.0%;  Score 38;  DB 12;  Length 1704;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-10;
  Matches   38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1025 gcccacctggacatctggaagtcagcggcctggatgaa 1062
          |||||||||||||||||||||||||||||||||||||
Db    941 GCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAA 978


RESULT   3
```

```
AF075706
LOCUS       AF075706       848 bp    mRNA            VRT       19-DEC-1998
DEFINITION  Gallus gallus ets domain protein (ERM) mRNA, partial cds.
ACCESSION   AF075706
NID         g3869357
VERSION     AF075706.1  GI:3869357
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 848)
  AUTHORS   Lin,J.H., Saito,T., Anderson,D.J., Lance-Jones,C., Jessell,T.M. and
            Arber,S.
  TITLE     Functionally related motor neuron pool and muscle sensory afferent
            subtypes defined by coordinate ETS gene expression
  JOURNAL   Cell 95 (3), 393-407 (1998)
  MEDLINE   99029846
REFERENCE   2  (bases 1 to 848)
  AUTHORS   Lin,J.H., Saito,T., Anderson,D.J., Lance-Jones,C., Jessell,T.M. and
            Arber,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1998) Center for Neurobiology & Behavior,
            Columbia University, 701 W. 168th St. HHSC 1013, New York, NY
            10032, USA
FEATURES             Location/Qualifiers
     source          1..848
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            <1..848
                     /gene="ERM"
     CDS             <1..712
                     /gene="ERM"
                     /codon_start=2
                     /product="ets domain protein"
                     /protein_id="AAC97201.1"
                     /db_xref="PID:g3869358"
                     /db_xref="GI:3869358"
                     /translation="RPQPAPPPPHGFQSPMGIKQEPPDYCIDSEVPNQQSSYVPGGLF
                     PSSHDGFSYEKDTRLYFDDTCVVPERLEGKVKQEPTLYREGPPYQRRGSLQLWQFLVT
                     LLDDPANAHFIAWTGRGMEFKLIEPEEVAPRWGIQKNPPAMNYDKLSRSLPYYYEKGI
                     MQKVAGERYVYKFVCDPDALFSMAYPDNQRPFLKAEPDCHVSEEDTLPLTHFEDSPAY
                     LLEVEHCGSLPYAEGFAY"
BASE COUNT      192 a    265 c    245 g    146 t
ORIGIN


  Query Match             1.6%;  Score 30;  DB 4;  Length 848;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-05;
  Matches    30;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1310 gccatgaactacgacaagctgagccgctcc 1339
           ||||||||||||||||||||||||||||||
Db     425 GCCATGAACTACGACAAGCTGAGCCGCTCC 454


RESULT   4
AF075708
LOCUS       AF075708      2508 bp    mRNA            VRT       19-DEC-1998
DEFINITION  Gallus gallus ets domain protein (PEA3) mRNA, complete cds.
ACCESSION   AF075708
NID         g4033764
VERSION     AF075708.1  GI:4033764
KEYWORDS    .
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
            Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2508)
  AUTHORS   Lin,J.H., Saito,T., Anderson,D.J., Lance-Jones,C., Jessell,T.M. and
            Arber,S.
  TITLE     Functionally related motor neuron pool and muscle sensory afferent
            subtypes defined by coordinate ETS gene expression
  JOURNAL   Cell 95 (3), 393-407 (1998)
  MEDLINE   99029846
REFERENCE   2  (bases 1 to 2508)
  AUTHORS   Lin,J.H., Saito,T., Anderson,D.J., Lance-Jones,C., Jessell,T.M. and
            Arber,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1998) Center for Neurobiology & Behavior,
            Columbia University, 701 W. 168th St. HHSC 1013, New York, NY
            10032, USA
COMMENT     On Dec 19, 1998 this sequence version replaced gi:3869361.
FEATURES             Location/Qualifiers
     source          1..2508
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
     gene            <1..2508
                     /gene="PEA3"
     CDS             <1..1011
                     /gene="PEA3"
                     /codon_start=1
                     /product="ets domain protein"
                     /protein_id="AAC97203.1"
                     /db_xref="PID:g4033765"
                     /db_xref="GI:4033765"
                     /translation="VHNNHRIVTLQYLFDSPSAYDQPRPVGIKSPDPGTPIQSPFQQF
                     PRQDRNFLTPSGPPHPTSSWGYLIENSSAYQAPVEMCHSFHSSQGMAQDAPIVYQRPM
                     SEQCLQCPPQGFKQEFHDPQYEQAGPSCQYPAAVVIKQFQVDYMYDSGCSYDKQMRPF
                     ADDVCVVPEKFEGDIKQEVGGYREGPPYQRRGSLQLWQFLVAILDDPTNSHFIAWTGR
                     GMEFKLIEPEEVAPLWGIQKNPPAMNYDKLSRSLPYYYEKGIMQKVAGFRYVYKFVCE
                     PEALFSLAFPDNQFPALKAELEEQISEFDTVPLSHLDESTAYIPLGSIPPQLYGKGY
                     TY"
BASE COUNT      563 a    715 c    672 g    557 t      1 others
ORIGIN


  Query Match             1.5%;  Score 29;  DB 4;  Length 2508;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-05;
  Matches    29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1310 gccatgaactacgacaagctgagccgctc 1338
           |||||||||||||||||||||||||||||
Db     724 GCCATGAACTACGACAAGCTGAGCCGCTC 752


RESULT   5
AC007818/c
LOCUS       AC007818    60056 bp    DNA             HTG       14-JUN-1999
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR02M05 (D785) RPCI-98
            02.M.5 map 98B-98B strain y; cn bw sp, WORKING DRAFT SEQUENCE, 20
            unordered pieces
ACCESSION   AC007818
NID         g5053146
VERSION     AC007818.1  GI:5053146
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 60056)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,F., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60056)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
```

```
            Doyle,C.M., Farfan,D.E., Galle,R., George,P.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.F., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,P., Zieran,L.I. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 400 bases, phrap computed error
            rate <= 1/1).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      893: contig of 893 bp in length
            *      894      973: gap of unknown length
            *      974     1964: contig of 991 bp in length
            *     1965     2044: gap of unknown length
            *     2045     3652: contig of 1608 bp in length
            *     3653     3732: gap of unknown length
            *     3733     5272: contig of 1540 bp in length
            *     5273     5352: gap of unknown length
            *     5353     6476: contig of 1124 bp in length
            *     6477     6556: gap of unknown length
            *     6557     8669: contig of 2113 bp in length
            *     8670     8749: gap of unknown length
            *     8750    11281: contig of 2532 bp in length
            *    11282    11361: gap of unknown length
            *    11362    13275: contig of 1914 bp in length
            *    13276    13355: gap of unknown length
            *    13356    16154: contig of 2799 bp in length
            *    16155    16234: gap of unknown length
            *    16235    18409: contig of 2175 bp in length
            *    18410    18489: gap of unknown length
            *    18490    20047: contig of 1558 bp in length
            *    20048    20127: gap of unknown length
            *    20128    21987: contig of 1860 bp in length
            *    21988    22067: gap of unknown length
            *    22068    24914: contig of 2847 bp in length
            *    24915    24994: gap of unknown length
            *    24995    28786: contig of 3792 bp in length
            *    28787    28866: gap of unknown length
            *    28867    32920: contig of 4054 bp in length
            *    32921    33000: gap of unknown length
            *    33001    36522: contig of 3522 bp in length
            *    36523    36602: gap of unknown length
            *    36603    41889: contig of 5287 bp in length
            *    41890    41969: gap of unknown length
            *    41970    46511: contig of 4542 bp in length
            *    46512    46591: gap of unknown length
            *    46592    51474: contig of 4883 bp in length
            *    51475    51554: gap of unknown length
            *    51555    60056: contig of 8502 bp in length.
FEATURES             Location/Qualifiers
     source          1..60056
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /clone="BACR02M05 (D785) RPCI-98 02.M 5"
                     /chromosome="3"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6"
                     /map="98B-98B"
BASE COUNT    15556 a  13397 c  13635 g  15945 t   1523 others
ORIGIN


  Query Match             1.5%;  Score 28;  DB 35;  Length 60056;
  Best Local Similarity   100.0%;  Pred. No. 0.00018;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1159 catccacctgtggcagttcctcaaggag 1186
           ||||||||||||||||||||||||||||
Db  54075 CATCCACCTGTGGCAGTTCCTCAAGGAG 54048


RESULT   6
DRODETS4A
LOCUS       DRODETS4A     362 bp    DNA             INV       05-AUG-1992
DEFINITION  Drosophila melanogaster Drets-4 DNA binding domain protein gene, 3'
            end.
ACCESSION   M88474
NID         g157195
VERSION     M88474.1  GI:157195
KEYWORDS    DNA-binding domain protein.
SOURCE      Drosophila melanogaster DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera, Endopterygota, Diptera, Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Chen,T., Bunting,M., Karim,F.D. and Thummel,C.S.
  TITLE     Isolation and characterization of five Drosophila genes that encode
            an ets-related DNA binding domain
  JOURNAL   Dev. Biol. 151, 176-191 (1992)
  MEDLINE   92249640
FEATURES             Location/Qualifiers
     source          1..362
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="Chromosome 3R 98B"
     gene            1..345
                     /gene="Ets98B"
                     /note="Ets at 98B"
                     /allele=""
                     /db_xref="FlyBase:FBgn0005659"
     CDS             1..345
                     /partial
                     /gene="Ets98B"
                     /codon_start=1
                     /product="D.melanogaster Drets-4 DNA binding domain protein"
                     /protein_id="AAA28451.1"
                     /db_xref="PID:g157196"
                     /db_xref="GI:157196"
                     /translation="TNASNGGTATVKPPNGGPIGGGGSHIHLWQFLKELLASPQVNGI
                     AIRWIDRSKGIFKIEDSVPVAKLWGPPKNPPAMNYDKLSRSIRQYYKKGIMKKTERSQ
                     RLVYQFCHPYSQ"
BASE COUNT       90 a     99 c    113 g     60 t
ORIGIN      Chromosome 3R, 98B.


  Query Match             1.5%;  Score 28;  DB 36;  Length 362;
  Best Local Similarity   100.0%;  Pred. No. 0.00017;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1159 catccacctgtggcagttcctcaaggag 1186
           ||||||||||||||||||||||||||||
Db     75 CATCCACCTGTGGCAGTTCCTCAAGGAG 102


RESULT   7
HSDNAFEV3
LOCUS       HSDNAFEV3    1403 bp    DNA             PRI       23-APR-1997
DEFINITION  H.sapiens FEV gene, exon 3.
ACCESSION   Y08979
```

```
NID          gi1841699
VERSION      Y08979.1  GI:1841699
KEYWORDS     ets gene family.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1403)
  AUTHORS    Peter,M., Couturier,J., Pacquement,H., Michon,J., Thomas,G.,
             Magdelenat,H. and Delattre,O.
  TITLE      A new member of the ETS family fused to EWS in Ewing tumors
  JOURNAL    Oncogene 14 (10), 1159-1164 (1997)
  MEDLINE    97246656
REFERENCE    2  (bases 1 to 1403)
  AUTHORS    Delattre,O.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-OCT-1996) O. Delattre, Institut Curie, Inserm U434,
             26 Rue Dulm, Paris 75231 Cedex 05, Paris 75231 Cedex 05, FRANCE
  REMARK     revised by EA
REFERENCE    3  (bases 1 to 1403)
  AUTHORS    Delattre,O.
  TITLE      Direct Submission
  JOURNAL    Submitted (14-FEB-1997) O. Delattre, Institut Curie, Inserm U434,
             26 Rue Dulm, Paris 75231 Cedex 05, Paris 75231 Cedex 05, FRANCE
COMMENT      On Feb 15, 1997 this sequence version replaced gi:1834485.
FEATURES             Location/Qualifiers
     source          1..1403
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /germline
     misc_feature    1..240
                     /number=2
                     /evidence=experimental
     gene            241..1403
                     /gene="FEV"
     exon            241..1403
                     /gene="FEV"
                     /note="ETS family member"
                     /number=3
                     /evidence=experimental
BASE COUNT      210 a    551 c    476 g    266 t
ORIGIN


  Query Match              1.4%;  Score 25;  DB 10;  Length 1403;
  Best Local Similarity   100.0%;  Pred. No. 0.0083;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1312 catgaactacgacaagctgagccgc 1336
           |||||||||||||||||||||||||
Db     422 CATGAACTACGACAAGCTGAGCCGC 446


RESULT   8
HSRNAFEV
LOCUS        HSRNAFEV     1901 bp    mRNA            PRI       23-APR-1997
DEFINITION   H.sapiens mRNA for FEV protein.
ACCESSION    Y08979
NID          gi1834508
VERSION      Y08979.1  GI:1834508
KEYWORDS     ets gene family.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1901)
  AUTHORS    Peter,M., Couturier,J., Pacquement,H., Michon,J., Thomas,G.,
             Magdelenat,H. and Delattre,O.
  TITLE      A new member of the ETS family fused to EWS in Ewing tumors
  JOURNAL    Oncogene 14 (10), 1159-1164 (1997)
  MEDLINE    97246656
REFERENCE    2  (bases 1 to 1901)
  AUTHORS    Delattre,O.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-OCT-1996) O. Delattre, Institut Curie, Inserm U434,
             26 Rue Dulm, Paris 75231 Cedex 05, Paris 75231 Cedex 05, FRANCE
FEATURES             Location/Qualifiers
     source          1..1901
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /cell_line="ewing"
     CDS             585..1301
                     /function="DNA binding protein"
                     /note="ETS family member"
                     /codon_start=1
                     /product="FEV protein"
                     /protein_id="CAA70169.1"
                     /db_xref="PID:e291921"
                     /db_xref="PID:g1834509"
                     /db_xref="GI:1834509"
                     /db_xref="SPTREMBL:Q99581"
                     /translation="MRQSGASQPLLINMYLPDPVGDGLFKDGKNPSWGPLSPAVQKGS
                     GQIQLWQFLLELLADRANAGCIAWEGGHGEFKLTDPDEVARRWGERKSKPNMNYDKLS
                     RALPYYYDKNIMSKVHGKRYAYRFDFQGLAQACQPPPAHAHAAAAAAAAAAAAQDGAL
                     YKLPAGLAPLPFPGLSKLNLMAASAGVAPAGPSYWPGPGPAATAAAATAALYPSPSLQ
                     PPPGPFGAVAAASHIGGHYH"
     misc_feature    711..962
                     /note="DNA binding domain"
BASE COUNT      302 a    707 c    575 g    317 t
ORIGIN


  Query Match              1.3%;  Score 25;  DB 10;  Length 1901;
  Best Local Similarity   100.0%;  Pred. No. 0.0083;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1312 catgaactacgacaagctgagccgc 1336
           |||||||||||||||||||||||||
Db     869 CATGAACTACGACAAGCTGAGCCGC 893


RESULT   9
DRODETS6A
LOCUS        DRODETS6A     480 bp    DNA             INV       05-AUG-1992
DEFINITION   Drosophila melanogaster D-ets-6 DNA binding domain protein gene,
             partial cds.
ACCESSION    M88475
NID          g157197
VERSION      M88475.1  GI:157197
KEYWORDS     DNA-binding domain protein.
SOURCE       Drosophila melanogaster DNA.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 480)
  AUTHORS    Chen,T., Bunting,M., Karim,F.D. and Thummel,C.S.
  TITLE      Isolation and characterization of five Drosophila genes that encode
             an ets-related DNA binding domain
  JOURNAL    Dev. Biol. 151, 176-191 (1992)
  MEDLINE    92249640
FEATURES             Location/Qualifiers
     source          1..480
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="Chromsome 2L 21C."
     intron          1..95
                     /gene="D-ets-6 DNA-binding domain"
     exon            96..288
                     /gene="Ets21C"
                     /db_xref="FlyBase:FBgn0005660"
     gene            join(96..288,366..480)
                     /gene="Ets21C"
                     /note="Ets at 21C"
                     /allele=""
```

```
                     /db_xref="FlyBase:FBgn0005660"
     CDS             join(96..288,366..480)
                     /partial
                     /gene="Ets21C"
                     /codon_start=3
                     /product="D-ets-6 DNA binding domain protein"
                     /protein_id="AAA28452.1"
                     /db_xref="PID:g157198"
                     /db_xref="GI:157198"
                     /translation="SGGQIQLWQFLLELLADSSNANAISWEGQSGEFRLIDPDEVARR
                     WGERKARPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGLMAAQQAQQGDP
                     "
     misc_feature    289..365
                     /gene="D-ets-6 DNA-binding domain"
     exon            289..480
                     /gene="Ets21C"
                     /db_xref="FlyBase:FBgn0005660"
BASE COUNT      131 a    138 c    142 g     99 t
ORIGIN      Chromosome 2L, 21C.


  Query Match              1.3%;  Score 24;  DB 36;  Length 480;
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1312 catgaactacgacaagctgagccg 1335
           ||||||||||||||||||||||||
Db     279 CATGAACTACGACAAGCTGAGCCG 256


RESULT 10
AC004573/c
LOCUS       AC004573    85095 bp    DNA             INV       30-JAN-1999
DEFINITION  Drosophila melanogaster, chromosome 2L, region 21C5-21D1, P1 clone
            DS07610, complete sequence.
ACCESSION   AC004573
NID         g4204255
VERSION     AC004573.1  GI:4204255
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 85095)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,P., George,R.A., Harris,N.I.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,P., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 2L, region 21C5-21D1
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 85095)
  AUTHORS   Celniker,S.E., George,R.A., Galle,P., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Iwomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,P.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Jan 30, 1999 this sequence version replaced gi:4164118
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu
            P1 library location: 80-26.
FEATURES              Location/Qualifiers
     source           1..85095
                      /organism="Drosophila melanogaster"
                      /strain="y2, cn bw sp"
                      /db_xref="taxon:7227"
                      /chromosome="2L"
                      /map="21C5-21D1"
                      /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
                      /clone="P1 DS07610 (D275)"
BASE COUNT    22289 a  19671 c  20319 g  22816 t
ORIGIN


  Query Match              1.3%;  Score 24;  DB 37;  Length 85095;
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1312 catgaactacgacaagctgagccg 1335
           ||||||||||||||||||||||||
Db    7600 CATGAACTACGACAAGCTGAGCCG 7577


RESULT 11
HS28C20/c
LOCUS       HS28C20    115835 bp    DNA             PRI       21-SEP-1998
DEFINITION  Human DNA sequence from clone 28C20 on chromosome 6q24. Contains a
            Zinc Finger gene ZNF131 pseudogene, ESTs an STS and GSSs, complete
            sequence.
ACCESSION   AL031119
NID         g3550036
VERSION     AL031119.1  GI:3550036
KEYWORDS    HTG; ZNF131.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 115835)
  AUTHORS   Smalley,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 8, 1998 this sequence version replaced gi:3355517.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 28C20. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed by the Sanger Centre Chromosome 6
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            28C20 is from the library RPCI1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.
FEATURES              Location/Qualifiers
     source           1..115835
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="6"
```

```
                     /clone="29528"
                     /map="q28"
                     /clone_lib="RPCI1"
repeat_region        1. .916
                     /note="L1 repeat: matches 3883. .4802 of consensus"
repeat_region        complement(1644. .2446)
                     /note="L1MB5 repeat: matches 870. .30 of consensus"
repeat_region        complement(2455. .2789)
                     /note="THE1C repeat: matches 356. .1 of consensus"
repeat_region        complement(2782. .2956)
                     /note="L1 repeat: matches 5284. .5111 of consensus"
repeat_region        3865. .4089
                     /note="MLT1F repeat: matches 317  .541 of consensus"
repeat_region        complement(8011. .8367)
                     /note="THE1B repeat: matches 364   6 of consensus"
repeat_region        complement(8641. .9132)
                     /note="L1MA9 repeat: matches 1017. .522 of consensus"
repeat_region        complement(9154. .10033)
                     /note="L1PA15 repeat: matches 904. .1 of consensus"
repeat_region        complement(9910   10804)
                     /note="L1 repeat: matches 5390. .4471 of consensus"
repeat_region        10805. .10838
                     /note="17 copies 2 mer gt 94% conserved"
gene                 complement(11946. .14372)
                     /gene="dJ28C20.1"
CDS                  complement(join(11946. .13396,14220. .14372))
                     /gene="dJ28C20.1"
                     /note="Zinc Finger gene ZNF131 pseudogene, match. proteins
                     P52739 Q99851 Q14526 P10074 Q15916 P24278 O35260 Q15062
                     Q92019 Q09376 O00403 Q13397 Q99592 P41183 P41182 Q13105
                     O73458 O15156 Q60821 O43298 P70616 Q03172 P28160 P15822
                     Q15918 O15044 Q05215 Q26452 Q24162 Q92981 O35745 P36508
                     P10925 Q60893 P18744 Q02313 Q02313 Q03924 Q14585 Q60493
                     Q14584 Q02403 Q08376 Q60821 O73453 Q62981 Q62513 O43693
                     Q14205 P51523 P52739; match: cDNA U09410; match: ESTs
                     F06005 R19845 R34147 F08171 AA482467 F12820 AA156200
                     AA482467 T75257"
                     /codon_start=1
                     /pseudo
                     /evidence=not_experimental
                     /db_xref="PID:g1321923"
repeat_region        complement(13829. .14060)
                     /note="LTR1 repeat: matches 224   1 of consensus"
repeat_region        complement(14650. .14787)
                     /note="LTR1 repeat: matches 674. .524 of consensus"
repeat_region        complement(15148. .15332)
                     /note="LTR1 repeat: matches 185. .1 of consensus"
repeat_region        complement(15504. .17889)
                     /note="L1 repeat: matches 4474. .2060 of consensus"
repeat_region        complement(17892. .18190)
                     /note="AluSq repeat: matches 300. .1 of consensus"
repeat_region        complement(18192. .19038)
                     /note="L1 repeat: matches 2071. .1248 of consensus"
repeat_region        complement(19045  .19263)
                     /note="AluYB8 repeat  matches 308   90 of consensus"
repeat_region        complement(19267. .19864)
                     /note="L1 repeat: matches 1250. .677 of consensus"
repeat_region        complement(21144. .21218)
                     /note="L1MA9 repeat: matches 529. .455 of consensus"
repeat_region        21810. .21956
                     /note="MIR3 repeat: matches 1   146 of consensus"
repeat_region        complement(23359. .23408)
                     /note="MIR repeat. matches 262. .216 of consensus"
repeat_region        complement(24208. .24359)
                     /note="MIR repeat: matches 156. .15 of consensus"
repeat_region        25546. .26822
                     /note="L1 repeat: matches 4106   5390 of consensus"
repeat_region        26673. .27565
                     /note="L1PA2 repeat: matches 1   893 of consensus"
repeat_region        complement(27574. .27920)
                     /note="MER42C repeat: matches 1536. .1189 of consensus"
repeat_region        28642. .28848
                     /note="L1ME1 repeat: matches 357. .561 of consensus"

repeat_region        complement(29341. .29382)
                     /note="AluSc repeat: matches 42. .1 of consensus"
misc_feature         30490. .30933
                     /note="match: genomic DNA Z78623"
repeat_region        30951. .31546
                     /note="MER25 repeat. matches 1515. .2124 of consensus"
repeat_region        31352   34748
                     /note="L1 repeat: matches 1431  .4813 of consensus"
repeat_region        34749. .35046
                     /note="AluSq repeat: matches 1. .300 of consensus"
repeat_region        35048   35628
                     /note="L1 repeat: matches 4796   5388 of consensus"
repeat_region        35484. .35901
                     /note="L1MA3 repeat: matches 1. .429 of consensus"
repeat_region        36692  .36867
                     /note="MLT2_internal repeat: matches 953. .1141 of
                     consensus"
repeat_region        36936. .37144
                     /note="MER20 repeat: matches 1   217 of consensus"
repeat_region        complement(37694. .38057)
                     /note="MLT1A2 repeat: matches 372. .1 of consensus"
repeat_region        38078. .38370
                     /note="MLT2_internal repeat: matches 1733. .2032 of
                     consensus"
repeat_region        38363   38592
                     /note="MLT2_internal repeat: matches 2221. .2461 of
                     consensus"
repeat_region        39064. .39677
                     /note="L1MB1 repeat: matches 289   916 of consensus"
repeat_region        40912. .40943
                     /note="16 copies 2 mer aa 97% conserved"
repeat_region        complement(44275. .44567)
                     /note="AluY repeat: matches 293. .1 of consensus"
repeat_region        45637. .46082
                     /note="TIGGER1 repeat: matches 25. .474 of consensus"
repeat_region        complement(46094. .46984)
                     /note="L1PA16 repeat: matches 904. .1 of consensus"
repeat_region        complement(46844. .50086)
                     /note="L1 repeat: matches 5390. .2153 of consensus"
repeat_region        50060. .52022
                     /note="TIGGER1 repeat  matches 430   2419 of consensus"
misc_feature         complement(52546. .53226)
                     /note="match: GSS B54551"
repeat_region        52809. .52983
                     /note="MIR repeat: matches 36   210 of consensus"
repeat_region        complement(57666. .57826)
                     /note="MER5B repeat: matches 178   3 of consensus"
repeat_region        58563. .58610
                     /note="24 copies 2 mer tt 81% conserved"
repeat_region        complement(58612. .58912)
                     /note="AluSx repeat: matches 300   1 of consensus"
misc_feature         59207. .59752
                     /note="match: GSS B57434"
repeat_region        complement(59910. .60566)
                     /note="L1PA8 repeat  matches 910. .256 of consensus"
misc_feature         60824. .61454
                     /note="match: GSS B99968"
repeat_region        63492. .63525
                     /note="17 copies 2 mer tg 97% conserved"
misc_feature         complement(64026. .64385)
                     /note="match: GSS AQ089270"
repeat_region        complement(65072. .65381)
                     /note="AluSp repeat. matches 303. .1 of consensus"
repeat_region        complement(65755. .65825)
                     /note="MIR repeat: matches 146   76 of consensus"
misc_feature         complement(66137. .66432)
                     /note="match: GSS AQ066992"
repeat_region        complement(68310. .68882)
                     /note="MLT2D repeat: matches 553   1 of consensus"
repeat_region        69187. .69291
                     /note="MIR repeat: matches 20   141 of consensus"
misc_feature         72197. .72557
                     /note="match: GSS B90081"
```

```
     repeat_region   72893. .73320
                     /note="MER28 repeat: matches 1. .434 of consensus"
     repeat_region   75287. .75648
                     /note="THE1D repeat: matches 1. .371 of consensus"
     repeat_region   complement(76395. .76458)
                     /note="MIR2 repeat: matches 134. .70 of consensus"
     repeat_region   complement(76658. .76719)
                     /note="MLT2C repeat: matches 388  324 of consensus"
     repeat_region   complement(76658. .76706)
                     /note="MLT2D repeat: matches 553. .505 of consensus"

  Query Match             1.2%;  Score 22;  DB 9;  Length 115835;
  Best Local Similarity   100.0%;  Pred. No. 0.4;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1525 ctgaactctgaccttccagaggcc 1546
           |||||||||||||||||||||||
Db   88123 CTGAACTCTGACCTTCCAGAGGCC 88102


RESULT  12
MMNEIRN
LOCUS       MMNEIRN      2658 bp    mRNA            ROD       13-MAR-1995
DEFINITION  M.musculus net mRNA.
ACCESSION   Z42815
NID         g479112
VERSION     Z42815.1  GI:479112
KEYWORDS    Net; ras gene.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2658)
  AUTHORS   Giovane A., Pintzas A., Maira S.M., Sobieszczuk,P. and Wasylyk,B.
  TITLE     Net, a negative factor switch to positive by Ras
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2658)
  AUTHORS   Giovane,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1994) Antoine Giovane,
            CNRS-LGME,INSERM-U.184,institut de Chimie, Biologique, 11 rue
            Humann, Strasbourg, 67085 Strasb. Cedex, France
REFERENCE   3  (bases 1 to 2658)
  AUTHORS   Giovane,A., Pintzas,A., Maira,S.M., Sobieszczuk,P. and Wasylyk,B.
  TITLE     Net, a new ets transcription factor that is activated by Ras
  JOURNAL   Genes Dev. 8 (12), 1502-1513 (1994)
  MEDLINE   95047810
FEATURES             Location/Qualifiers
     source          1. .2658
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /dev_stage="10 days"
                     /tissue_type="embryo"
                     /clone_lib="lambda zap2"
     CDS             295. .1524
                     /codon_start=1
                     /product="Net"
                     /protein_id="CAA83676.1"
                     /db_xref="PID:g479113"
                     /db_xref="GI:479113"
                     /db_xref="MGD:MGI:101762"
                     /db_xref="SWISS-PROT:P41971"
                     /translation="MESAITLWQFLLHLLLDQKHEHLICWTSNDGEFKLLKAEEVAKL
                     WGLRKNKTNMNYDKLSRALRYYYDKNIIKKVIGQKFVYKFVSFPDILKMDPHAVEISR
                     ESLLLQDGDCKVSPEGPEVHPHGLSSLKSASPNEYLHSGLYSSFTINSLQNAPEAFKA
                     IKTEKLEEPCDDSPPVEEVPTVIPFVTNKTDKHITPPVMSLPSTSETAAAAASAFLAS
                     SVSAKISSLMLPNAASVSSVSPSSSRSPSLSPDSPLPSEHRSLFLEAACHESDSLEPL
                     NLSSGSKTKSPSLPPKGKKPKGLEISAPQLLLSGTDIGSIALNSPALPSGSLTPAFFT
                     AQTPSGLFLASSPLLPSIHFWSSLSPVAPLSPARLQGPNTLFQFPTLLNGHMPVPLPS
                     LDRAPSPVLLSPSSQKS"
BASE COUNT      682 a    714 c    600 g    662 t
ORIGIN
```

```
  Query Match             1.2%;  Score 22;  DB 12;  Length 2658;
  Best Local Similarity   100.0%;  Pred. No. 0.39;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1312 catgaactacgacaagctgagc 1333
          |||||||||||||||||||||||
Db    453 CATGAACTACGACAAGCTGAGC 474


RESULT  13
MUSTRANONC
LOCUS       MUSTRANONC   2211 bp    mRNA            ROD       19-JUL-1994
DEFINITION  Mus musculus ERP mRNA, complete cds.
ACCESSION   L19953
NID         g436184
VERSION     L19953.1  GI:436184
KEYWORDS    transcription factor.
SOURCE      Mus musculus (strain BALB/c, sub_species domesticus) (library:
            lambda-gt11 of Clontech) male lung cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2211)
  AUTHORS   Lopez,M., Oettgen,P., Akbarali,Y., Dendorfer,U. and Libermann,T.A.
  TITLE     ERP, a new member of the ets transcription factor/oncoprotein
            family, cloning, characterization, and differential expression
            during B-lymphocyte development
  JOURNAL   Mol. Cell. Biol. 14 (5), 3292-3309 (1994)
  MEDLINE   94217726
FEATURES             Location/Qualifiers
     source          1. .2211
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /sub_species="domesticus"
                     /db_xref="taxon:10090"
                     /sex="male"
                     /tissue_type="lung"
                     /tissue_lib="lambda-gt11 of Clontech"
     gene            302. .2199
                     /gene="ERP"
     CDS             302. .1531
                     /gene="ERP"
                     /codon_start=1
                     /protein_id="AAA19872.1"
                     /db_xref="PID:g436185"
                     /db_xref="GI:436185"
                     /translation="MESAITLWQFLLHLLLDQKHEHLICWTSNDGEFKLLKAEEVAKL
                     WGLRKNKTNMNYDKLSRALRYYYDKNIIKKVIGQKFVYKFVSFPDILKMDPHAVEISR
                     ESLLLQDGDCKVSPEGPEVHPHGLSSLKSASPNEYLHSGLYSSFTINSLQNAPEAFKA
                     IKTEKLEEPCDDSPPVEFVPTVIPFVTNKTDKHITPPVMSLPSTSETAAAAASAFLAS
                     SVSAKISSLMLPNAASVSSASPSSSFSPSLSPDSPLPSEHRSLFLEAACHESDSLEPL
                     NLSSGSKTKSPSLPPKGKKPKGLEISAPQLLLSGTDIGSIALNSPALPSGSLTPAFFT
                     AQTPSGLFLASSPLLPSIHFWSSLSPVAPLSPARLQGPNTLFQFPTLLNGHMPVPLPS
                     LDRAPSPVLLSPSSQKS"
     misc_feature    320. .553
                     /gene="ERP"
                     /note="ETS domain"
     polyA_signal    2195. .2199
                     /gene="ERP"
BASE COUNT      560 a    613 c    504 g    534 t
ORIGIN


  Query Match             1.2%;  Score 22;  DB 12;  Length 2211;
  Best Local Similarity   100.0%;  Pred. No. 0.39;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1312 catgaactacgacaagctgagc 1333
          |||||||||||||||||||||||
Db    460 CATGAACTACGACAAGCTGAGC 481
```

```
RESULT  24
S82864
LOCUS       S82864          1526 bp    mRNA            ROD       10-FEB-1997
DEFINITION  Elk-3 Ets transcription factor [mice, 16-day embryos, mRNA, 1526
            nt].
ACCESSION   S82864
NID         g1836129
VERSION     S82864.1  GI:1836129
KEYWORDS    .
SOURCE      Mus sp. 16-day embryos.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1526)
  AUTHORS   Nozaki,M., Onishi,Y., Kanno,N., Ono,Y. and Fujimura,Y.
  TITLE     Molecular cloning of Elk-3, a new member of the Ets family
            expressed during mouse embryogenesis and analysis of its
            transcriptional repression activity
  JOURNAL   DNA Cell Biol. 15 (10), 855-862 (1996)
  MEDLINE   97047916
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 179297] from the original journal article.
            This sequence comes from Fig. 1.
FEATURES             Location/Qualifiers
     source          1..1526
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
     gene            94..1323
                     /gene="Elk-3"
     CDS             94..1323
                     /gene="Elk-3"
                     /note="Ets transcription factor; This sequence comes from
                     Fig. 1"
                     /codon_start=1
                     /product="Elk-3"
                     /protein_id="AAB46893.1"
                     /db_xref="PID:g1836130"
                     /db_xref="GI:1836130"
                     /translation="MESAITLWQFLLHLLLDQKHEHLICWTSNDGEFKLLKAEEVAKL
                     WGLRKNKINMNYDKLSRALRYYYDKNIIKKVIGQKFVYKFVSFPDILKMDPHAVEISR
                     ESLLLQDGDCKVSPEGREVHRHGLSSLKSASRNEYLHSGLYSSFTINSLQNAPEAFKA
                     IKIEKLEFPCQDSPPVEEVRTVIRFVTNKTDKHITPPVVSLPSTSETAAAAASAFLAS
                     SVSAKISSIMLPNAASVSSASPSSSRSPSLSPDSPLPSEHRSLFLFAACHESDSLEPL
                     NLSSQSKIKSPSLPPKGKKPKGLEISAPQLLLSGTDIGSIALNSPALPSGSLTPIFFT
                     AQIPSGLELASSPIIPSIHFWSSISPVAPLSPARLQGPNTLFQFPTLINGHMPVPIPS
                     LDRAPSPVLLSPSSQKS"
BASE COUNT       453 a    487 c    350 g    336 t
ORIGIN


  Query Match             1.2%;  Score 22;  DB 12;  Length 1526;
  Best Local Similarity   100.0%;  Prod. No. 0.39;
  Matches   22;  Conservative    0,  Mismatches    0,  Indels    0,  Gaps    0.

Qy     141 catgaactacgacaagcigagc 183
           |||||||||||||||||||||
Db     252 CATGAACTACGACAAGCTGAGC 273


RESULT  25
PFMAL4P4/c
LOCUS       PFMAL4P4    113899 bp    DNA             INV       29-APR-1999
DEFINITION  Plasmodium falciparum MAL3P4, complete sequence.
ACCESSION   AL008970
NID         g4493924
VERSION     AL008970.2  GI:4493924
KEYWORDS    HTG.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 113899)
  AUTHORS   Hamlin,N., Bowman,S., Churcher,C., Lawson,D., Quail,M. and
            Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Mar 24, 1999 this sequence version replaced gi:3763998.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is not the entire insert of clone MAL3P4. It may be
            shorter because we only sequence overlapping sections once, or
            longer because we arrange for a small overlap between neighbouring
            submissions.
FEATURES             Location/Qualifiers
     source          1..113899
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="3"
                     /clone="MAL3P4"
     gene            2364..6248
                     /gene="MAL3P4.26"
     CDS             2364..6248
                     /gene="MAL3P4.26"
                     /note="predicted using hexExon; MAL3P4.26 (PFC0435w),
                     Hypothetical protein, len: 1295 aa"
                     /codon_start=1
                     /protein_id="CAA15616.1"
                     /db_xref="PID:e1332567"
                     /db_xref="PID:g3764023"
                     /db_xref="GI:3764023"
                     /db_xref="SPTREMBL:O77361"
                     /translation="MKVGIIFFCLFFFVVLGACNNVKERIFKNIKKRTKFIILNEPIV
                     DLSFSENLFHTLLFDLDVDKNLYTLDESLLNLENLNYSSIFPLLVDIYKNIKENEDDN
                     KNIRYIFLGTSFSRIHPLNFEYFLRKLNKYIYNGNIYEKGNVDIRGILEEYNKEIEEK
                     KLEKQKLNKIKDKNNNNNNNNNSKFSKDGDNEDFNNKNDLYNPSDKLYNNNDDIDVHE
                     LLEEIITKEKREFLNDDDDNDSNDKYILKIDEVNKYKGFFIGYGFNDDIPSVIHHYNF
                     DKNFLFPSLNSGIILDITLLKNIYEVSNILLSNNEKDQSHIDYIYEVTKYIKFNLPV
                     PLTHSENVCLNEEQNIHLLDNDPNNFEIYKYYQVLNLFKDYNKNTFEKQYEKIGHENV
                     PHEETSSEGNENLNFNTKHNNDNNNDNNNYSEDAIAELLLSYFNVFYPISICMCYSIP
                     SKHESLMDYDKYHMINLENDIKLKHYIKETEEIHFNSIEEYKMKLNRINYKYDTLLEE
                     HENLVTHKNILIGIKTSINTEEERIPHIKNTYDNKENTQIIFNTFNYDNKLKEKNTFG
                     FYNNSLLQNALENDNIDLDIIYMSDKFSQKYDNLYFNSKVTSKFGLCFKLKHMIYYYY
                     EEYVMKNSEKKYFFIADDDTFVNVKNLIDVTNLTLNTCSHSKKYMYDKYIKSYDFVKE
                     NEALFLQNFPKKILFLYSYLKDIFAKIIQILKKYDYVPKYCQGGILSKKHKNNDSDDD
                     HDHHVGNKQNNDSINHQDIEKNQVNVINNNNNNNNNKAKSIPIYLGKRYSYNTFSTNS
                     NEYFYDYLTGGAGILINDETAKPIYECKFCTCPSTNSSMDDMIPGKWAKELGILAINF
                     EGYFQNSPLDYNKKYINTLVPITYHPLNKNRTTKESRDMYFNYLVNYNRNDKEQNKDI
                     YVDYLDRNHKNMIDNVFHYFFYVNMYDEKNKVVTKIEHNADMNSKKNKSKNPQKLNNT
                     QGDKNVNDDENVNDDENVKGDENVKGDEYMKGDFNVKGDENVKDDENVKDDE
                     NIKGDDNNYNVDNMENIDDIINMVESVDDDVMERNKKGTGKEKKDDKNHNNKEKATDV
                     KKSSVPTNNIDKNEDTTKYVIKMNEKIYNRMQESGKYKQLFDINKFFKKEIEGHPYFQ
                     KIKKKNEKAKKEKFKMNQLKKQKDYTNNYFHTSNMQGNFNQQFMGNYGNQENEENDFF
                     DQPPEIEFDAINPMDYEEYMFNLSNFFDDGEPYDFYDDYDDFVNTINADKIKINDQNK
                     HLYEQIKDIAQPPVNFQNDQNSNTFDFDTDEL"
     gene            complement(7161..13970)
                     /gene="MAL3P4.25"
     CDS             complement(7161..13970)
                     /gene="MAL3P4.25"
                     /note="predicted using hexExon; MAL3P4.25 (PFC0440c),
                     Helicase, len: 2270 aa; Similarity to helicases. C.elegans
                     helicase (WP:F52B5.3) BLAST Score: 290, sum P(4) =
                     5.9e-31; 34% in 199 aa overlap."
                     /codon_start=1
                     /protein_id="CAA15615.1"
                     /db_xref="PID:e1332566"
                     /db_xref="PID:g3764022"
                     /db_xref="GI:3764022"
                     /db_xref="SPTREMBL:O77360"
                     /translation="MNNTFKINKKRPTVYFSYNIINIANKKRYMCNDNKMDNDNNIIN
                     EKKYFNNSSSIKNTGNNYKNDNIDKSYDMCERRSMMDRDENIYNNNYKNKNRNNYSNK
                     NNYSNNINHMDNRRNTYKGDLDGRKQYLSNKHYESKNDNENSYNLKKYNFNIPPIKYN
                     NNNNNNNNNNNNDRQNIDNNIYKFYDDKNEKKSNKTHSYGNNFNNEFNTFNDPKQNYM
                     SNESHNMIQDGHPIYNHNNNNNDGKIKNKKDISLSSYGYNKISNDNISKENTSYINQY
                     NNSGGANRHVINAQHNNNQHNNNQHNNNFNYHNYKNAKPYQDKYKNGDSFKFSAYNAE
```

```
                     QHDNMDKNYYQDIHKNNMKHSTKGDIAISFNGQKVFGSFKNKQMKEKRKFFKNKERGS
                     SINRPGMHNNNNNYNNNNNNNYYNNYNNHFCDDAERRNNKNFNWCFINSNEIYEKDGD
                     MPIYNANENCNLKKMIYKNDKKEKEKIKILLTEDNVNNVQSNIEKLSIYKSRNEIIEM
                     IEKNDVTFINGFTGSGKSICVPKFLLEENIPENKKINIIVTFPRRIACIALSKILSEL
                     INEKIGLKIRYRIGGFSLYDSEKTVITYITIGYLFKLFLHHKNMYKKFTHVIIDEIHE
                     RSILLDIVLLFIKLYLHNKQKDEQMFKLIIMSATMQSNLFYSYFEHPNIKMGSIFIGT
                     KIFSIDIFYIEDIINYIPYGSRNIHDEKINDEKINDEKINDEKINDDNKKSCD
                     DKKSTDAKSCDGNNNFVKFSVIDFILRPKNCNKINLSKNSFMLLYKIKSEYDKNIHM
                     FNNNNNKYCNDKNKDLDVDEIIPANVFSNISNLCLELVYNLCLKGDSVIIFLSGMQDI
                     IDMYHQISMIINNNMDNPSNSNININNGNDANDANDVNNMHNMHNMHNIANQLTY
                     HRTDIKIHIHMLHSCIYDNTIHKLKHNDTDINIFLSSNIAESSITIPNVPLVIDFCIQ
                     KNIFYNQKKKAHILVKKWINKSSMEQRKGRCGRTCHGICIPMISKNFLNLLPDHKISE
                     IYTHSLHLLYLYILKSMSVLNGLINKRNEVVHCDNTNVYNVKEVKDQKNGICKNEISH
                     SHAFNKKITIYDVLSMIIEKPSKEKIKSTPYELEKVKAVIKIKDKLVISIIQQIMIPF
                     NLSINLCRLLLYGVLLDVTFDTIILIGILNTNDIFPNINLYSSKNIYSYAVSLEICSK
                     QKAYFDGNTYSEPIMLPNVFLEWLCVYLLYVQGLKKENKFHPKELKTYYMNTCSIMNK
                     RNHVNSKKLLCVINSVDNLCKKMLKMLNKNSNAYKSCVYLLYLLPGGSDINYNMGNNS
                     NNNNDNSGEIINADGITPNNTHMNEINGMNVFNVVTKYCLFDYSNQNIYLKFLFSLSF
                     TPLFIHGSPNLPLRDLNEGKKKSKKLMNVLNFMIDNKLDVKNCIYFSGVYIPDINILK
                     RAICIMCPYISFDIYVSKNIYYVYFHNNNYHNNYHNNILKGDDFSYYNKTICDKGDAD
                     MSVNSYLNLTQEGHDTDGITSFGVSPLMNNNNMYGNNYIHNNVYYPCNGINNYYNNNV
                     VNSQNNYMMKMIMSNYYSFSFLNLTSGKNYPDIIEKIRNKLIYKYNNNVEKSIFHNFY
                     NNILNLIQPVNAIVPVYSNKYDEINNASICTNIINMFSNGKWIFTMPLFNPNKRKCEV
                     SEYFNHKYFLKKPKHLFLVKWMFLNTDNYKIKKNKQKEHVMDTNNALVNNNSLGKNVS
                     KNCIRFDNINLNYSLNDQEDQQYNKGYYYDDYDDYDDYDDDDNNDDNNDNNNNNDDNN
                     NNNDDDYLSNVKVESDNNNTHQNNPDGYEGNDSSSMCSDNNKKSKQKKMKCNLNFRSV
                     LGFLSMCPFNFDTKRNIYDNQTADIYAVCSSIDYSYTNDTYTWVNYVTVLPNKYFLTF
                     FLSSIPYHDNVIIQTRTNLYTSDILSIKIFDSKEIILKNMKKMKHKNKTNTTNYYDTN
                     STINNPNITKHDLLRVNYLRYVMSKLLLSLTLSFNPQEIEQNTQPLKLSEKSKPKSVI
                     INKIWRPNNDNRVDANDKNVNGNDKNVNDNDKNVNSQDNQGNQKDHFDDEDEEYFSDED
                     IMSDIEENNILKENLYLKQIVINMIEENNMDYINYNKYSMDNMDIDEHFNLQYLYEDE
                     HFINLNEKWYNNKGEILMNTYDYLYNSLKNENDYNQYNNNNNNNNNNNNNSSYMVKNI
                     KFKAINSIFAVKSSYINNNFSVQVQNTMNAYERTNSEDFLFFQPINISPIKEANYKLF
                     SMYYNEVAPTLDKFRRTNRQV"
gene                 18825. .19157
                     /gene="MAL3P4.24"
CDS                  18825. .19157
                     /gene="MAL3P4.24"
                     /note="predicted using hexExon; MAL3P4.24 (PFC0445w),
                     Hypothetical protein, len: 111 aa"
                     /codon_start=1
                     /protein_id="CAA15613.1"
                     /db_xref="PID:e1332564"
                     /db_xref="PID:g3764020"
                     /db_xref="GI:3764020"
                     /db_xref="SPTREMBL:O77358"
                     /translation="MLHGISTISEKINVYSLYEKKTNIFQSLEKKGIETIEGDGFKIQ
                     CYDTLTGIKIFIVHKDDLNIEMNTYLKPVYELYSDIILKNPFYDIDMPIPSAVFNEQI
                     EKLFSNIS"
gene                 19545. .19874
                     /gene="MAL3P4.23"
CDS                  19545. .19874
                     /gene="MAL3P4.23"
                     /note="predicted using hexExon; MAL3P4.24 (PFC0450w),
                     Hypothetical protein, len: 110 aa"
                     /codon_start=1
                     /protein_id="CAA15614.1"
                     /db_xref="PID:e1332565"
                     /db_xref="PID:g3764021"
                     /db_xref="GI:3764021"
                     /db_xref="SPTREMBL:O77359"
                     /translation="MSMFLNILILIDAASVAFLLITFLMINLNEESLELSQAHPENGK
                     KAIVVAIIYVIFIVILFIYKAYKNKPKLYTNFFMKKPNAPKYVQLASTYLSASDEYE
                     QYELNKI"
gene                 20968. .23328
                     /gene="MAL3P4.22"
CDS                  20968. .23328
                     /gene="MAL3P4.22"
                     /note="predicted using hexExon, MAL3P4.22 (PFC0455w),
                     Hypothetical protein, len: 787 aa"
                     /codon_start=1
                     /protein_id="CAA15612.1"
                     /db_xref="PID:e1332563"
                     /db_xref="PID:g3764019"
                     /db_xref="GI:3764019"
                     /db_xref="SPTREMBL:O77357"
                     /translation="MKLLKLDYYNGEIYEKYKINNILSLLVDNNKFQNESVDFHEISN
                     KKNNYSIMAILSSQIKNLNDDIYKDIDINDIILALSKLKFIDEEILINESQNKNYILE
                     LAEQIYYEYFQRGLIKNDHNFHLNNSKKNFIQINIELMNEINREIKNYEHKIHNLRAY
                     SENAQTKNDASINDLLMYYHEDVDNIDDSLLNENNFNYNMKHKKYNDQNKIFLNDHAH
                     NNNHNNNNNDFFDRPNENNFLMDEENVLKAAHEDIPMNNDYVNYIHMKRDSLHNNNNN
                     NNNNNYNNNYNNYKNNIFSDHHDLFHYNKEGTNSTNCNSVENLNYDINIYKNNKKNI
                     QHINNNVNSNENIPPVSNPNGNIGNDGPNSYPNLSQLNEEDKHILYNDTIKMDYILNP
                     KIEYEKLDEDQSESNNKSMMVMNTHPMMKNFDNMNKINSNVVDTYEENMNYKYDKNKS
                     YVLNDEETYEHMSNKNMNNVMKNTGANNGANNGVEENCKYHIDMDNVHNAVNSVNLQD
                     NIYKNNNSSNTLNSVLENIDKAEQQIVKNHLKNSLAHNEYVAKRMSSLLLNPQNVIDK
                     EPYNNNIPKAYEKNGIMNPDPQKYFNDFISLNKYSSIYGQKKNFTIDTNNSSSFILST
                     IHKPNMVDESSYNIMNNYENKKFRIQSRNMNTRENYLHNQNSNHHSTERSHNINNNMI
                     NHNNHISNTTPLNFDNMSNYSPNAAKSVESNMNIYTNKYSPALDINDLNSNNISPNYV
                     NSSNSLKSDLTMMRDNYAKMLERSKSLTSHLYRAPIMHNKYLKANM"
gene                 24170. .25001
                     /gene="MAL3P4.21"
CDS                  join(24170. .24420,24635. .25001)
                     /gene="MAL3P4.21"
                     /note="predicted using hexExon; MAL3P4.21 (PFC0460w),
                     Hypothetical protein, len: 206 aa"
                     /codon_start=1
                     /protein_id="CAA15611.1"
                     /db_xref="PID:e1332562"
                     /db_xref="PID:g3764018"
                     /db_xref="GI:3764018"
                     /db_xref="SPTREMBL:O77356"
                     /translation="MYTYITYLNIYFIKIPMMMNPTLGKPEEDLAVYDYTDISMNVCL
                     KKYEYIFDPWTNDSRYLYSNENMQNEIEALNELMKSKVSLVCAFFNPAFSYLFMTIGF
                     IGMFVFYKMKNDLSMGFIFACIFLVSFSHYIMCIICSYIITKDVIKSVETPIKSLNIK
                     YKKEKIYFKLINLSSFGIKYLKLFTLSKGQAYFTIPVIYKYCKII"
gene                 complement(26011. .28044)
                     /gene="MAL3P4.20"
CDS                  complement(join(26011  27209,27372  27599,27795  28044))
                     /gene="MAL3P4.20"
                     /note="predicted using hexExon, MAL3P4.20 (PFC0465c),
                     Hypothetical protein, len. 559 aa"
                     /codon_start=1
                     /protein_id="CAA15610.1"
                     /db_xref="PID:e1332561"

  Query Match             1.2%,  Score 22;  DB 36;  Length 113899;
  Best Local Similarity   100.0%;  Pred. No. 0.4;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1232 aacaaggagaaggcatcttca 1253
          ||||||||||||||||||||||
Db  90303 AACAAGGAGAAGGGCATCTTCA 90282


Search completed: November 20, 1999, 15:27:56
Job time: 669 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 14:10:13 ; Search time 11.57 Seconds
(without alignments)
1160.064 Million cell updates/sec

Title: US-09-126-945-2
Perfect score: 1785
Sequence: 1 MGSASPGISSVSPSHLLLPP.....GIIRKPDISQRLVYQFVHPI 335

Scoring table: BLOSUM62

Searched: 122810 seqs, 40055486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	354	19.8	114	2	S28821	transcription fact
2	302.5	16.9	486	1	TVHUEG	transforming prote
3	294	16.5	464	1	S37616	transcription fact
4	293	16.4	454	2	I38739	nuclear respirator
5	292	16.4	479	1	S60754	transcription fact
6	289	16.2	454	1	A48146	GA-binding protein
7	277	15.5	454	1	A40858	GA-binding protein
8	275	15.4	453	1	S49013	transforming prote
9	272	15.2	485	1	TVCHET	transcription fact
10	272	15.2	452	1	S17403	transforming prote
11	270	15.1	452	1	I37565	transforming prote
12	266	14.9	488	1	TVFVES	transcription fact
13	257	14.4	581	2	G02318	Ets transcription
14	256.5	14.4	300	2	JW0048	ets transcription
15	254.5	14.3	441	1	TVCHTE	transcription fact
16	253.5	14.2	807	2	A53225	ecdysone-induced p
17	253.5	14.2	972	2	B53225	ecdysone-induced p
18	252.5	14.1	929	2	A34692	ecdysone-induced p
19	252.5	14.1	993	2	B34692	ecdysone-induced p
20	250	14.0	472	1	A53236	transcription fact
21	244.5	13.7	385	1	S29844	transforming prote
22	247.5	13.9	983	2	S04722	puff 74E protein -
23	246.5	13.8	469	1	TVHUE2	transcription fact
24	246	13.8	468	1	TVMSE2	transcription fact
25	244.5	13.7	438	1	S11225	transcription fact
26	243.5	13.6	440	1	A35875	transcription fact
27	243	13.6	472	1	B53236	transcription fact
28	241.5	13.5	441	1	TVHUET	transcription fact
29	240.5	13.5	441	1	A53988	transcription fact
30	237	13.3	479	1	TVCHE2	transcription fact
31	236	13.2	510	1	S43692	transcription fact
32	233.5	13.1	619	2	A43361	Ets-related transc
33	221	12.5	555	1	S24061	transcription fact
34	224	12.5	428	1	TVHUEK	transforming prote
35	224	12.5	429	2	JC4965	elk1 protein - mou
36	222.5	12.5	722	2	A43315	ETS domain protein
37	222	12.4	462	1	S35524	adenovirus E1A enh
38	218.5	12.2	440	2	I48291	transcription fact
39	217.5	12.2	477	1	I38993	transcription fact
40	216.5	12.1	453	2	B42093	serum response fac
41	216.5	12.1	405	2	A42093	serum response fac
42	212.5	11.9	477	1	B46396	transcription fact
43	210.5	11.8	430	2	I48755	mSAP1a - mouse
44	208	11.7	761	2	A46193	88K E-26-specific
45	205.5	11.5	623	1	S33167	pointed protein, s

ALIGNMENTS

RESULT 1
S28821
transcription factor ets-98B - fruit fly (Drosophila melanogaster) (fragment)
N;Alternate names: transforming protein ets-4
C;Species: Drosophila melanogaster
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 24-Sep-1998
C;Accession: S28821
R;Chen, T.; Bunting, M.; Karim, F.D.; Thummel, C.S.
Dev. Biol. 151, 176-191, 1992
A;Title: Isolation and characterization of five Drosophila genes that encode an ets-r
A;Reference number: S28819; MUID:92249640
A;Accession: S28821
A;Molecule type: DNA
A;Residues: 1-114 <CHE>
A;Cross-references: EMBL:M88474, NID:g157195, PID:g157196
C;Genetics:
A;Gene: Ets98B
A;Cross-references: FlyBase:FBgn0005659
C;Superfamily: ets DNA-binding domain homology
C;Keywords: DNA binding; nucleus; transcription factor
F;28-109/Domain: ets DNA-binding domain homology <ETS>

Query Match 19.8%; Score 354; DB 2; Length 114;
Best Local Similarity 74.2%; Pred. No. 1.2e-22;
Matches 66; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

```
Qy     246 GQPIHLWQFLKELLLKPHSYGPFIPWLNKEKGIFKIEDSAQVAPLWGIFKNPPAMNYDKL 305
           |  ||||||||||  |   |  |||::: |||||||| :||:||| ||||||||||||
Db      23 GSHIHLWQFLKELLASPQVNGIAIRWIDRSKGIFKIEDSVPVAKLWGFFKNPPAMNYDKL 82

Qy     306 SRSIRQYYKKGIIRKPDISQRLVYQFVHP 334
           ||||||||||||::| : |||||||| ||
Db      83 SPSIPQYYKKGIMKKTEPSQRLVYQFCHP 111
```

RESULT 2
TVHUEG
transforming protein erg-3 - human
N;Contains: transforming protein erg-1; transforming protein erg-2
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1989 #sequence_revision 30-May-1997 #text_change 31-Oct-1997
C;Accession: A94294, A94178, I58410, A28041, A29515
R;Rao, V.N.; Papas, T.S.; Shyam, P.; Reddy, P.
Science 237, 635-639, 1987
A;Title: erg, a human ets-related gene on chromosome 21: alternative splicing, polyad
A;Reference number: A94294; MUID:87263429
A;Accession: A94294
A;Molecule type: mRNA
A;Residues: 1-231,256-486 <REA>
A;Cross-references: GB:M17254; NID:g182186; PID:g182187
R;Reddy, E.S.P.; Rao, V.N.; Papas, T.S.
Proc. Natl. Acad. Sci. U.S.A. 84, 6131-6135, 1987
A;Title: The erg gene: a human gene related to the ets oncogene.
A;Reference number: A94178; MUID:87317608
A;Accession: A94178
A;Molecule type: mRNA
A;Residues: 100-231,256-486 <REB>
A;Cross-references: GB:M21535; NID:g182182; PID:g182185
R;Prasad, D.D.; Rao, V.N.; Lee, L.; Reddy, E.S.
Oncogene 9, 669-673, 1994
A;Title: Differentially spliced erg-3 product functions as a transcriptional activato

A;Reference number: I58410; MUID:94119611
A;Accession: I58410
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 240-259 <RES>
A;Cross-references: GB:S68130; NID:g544960; PID:g544961
C;Genetics:
A;Gene: GDB:ERG
A;Cross-references: GDB:119884; OMIM:165080
A;Map position: 21q22.2-q22.2
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription fa
F;1-241,256-486/Product: transforming protein erg-2 #status predicted <EG2>
F;100-241,256-486/Product: transforming protein erg-1 #status predicted <EG1>
F;126-206/Domain: ets RII regulatory region homology <ETS2>
F;320-398/Domain: ets DNA-binding domain homology <ETS>

Query Match 16.0%; Score 302.5; DB 1; Length 486;
Best Local Similarity 26.0%; Pred. No. 1.6e-17;
Matches 106; Conservative 45; Mismatches 121; Indels 135; Gaps 16;

```
Qy      35 PSPPAIPEQGLSAE--------------YLSYFDML------YPEDSSWAAKAPGASSPE 84
           | | :  :: ||                :|:  :|       | : |  : : |
Db       6 PDPAAHIKEALSVVSEDQSLFECAYGTPHLAKTEMTASSSSDYGQTSKMSPRVPQQDWLS 65

Qy      85 EPP-  -  --------------EEPEQCPVIDSQAPAGSLDLVP---GGLTLEEHSLEQVQ 123
           :||          |:.| |          || | |   |  |.|
Db      66 QPPARVTIKMECNPSQVNGSRNSPDECSVAKGGKMVGSPDTVGMNYGSYMEEKHM--PPP 123

Qy     124 SMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLLWTEHQYRLPPMG-KAFQELAGKE 182
           :|   |          : : : |||  ||  :|::|| |   :| || :    || : |||
Db     124 NMITNE--      ---RRVIVPADPTLWSTDHVRQWLEWAVKEYGLPDVNILLFQNIDGKE 174

Qy     183 LCAMSEEQFRQRSP-LGGDVLHAHL--------------DIWK----SAAWMKERTSPGA 223
           :| |:::  |:: :|    |:| :||              |: |    |   |  | : ||
Db     175 LCKMTKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTGGA 234

Qy     224 IHYCASIS--------------------EESW-------------------------- 236
              :|'                      :|
Db     235 AFIFPNISVYPEATQRITTRPDLPYEPPRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQR 294

Qy     237      -  -    -TDSPVDSSCSSQPIHLWQFLKELLLKPHSYGRFIRWLNKEKGIFKIEDS 284
                  | | :  ||| | |||| | ||   |    | |      | ||: |
Db     295 PQLDPYQILGPTSSRLANPGSGQ-IQLWQFLLE-LLSDSSNSSCITW-EGTNGEFKMTDP 351

Qy     285 A-VARLW-I-R-SPFAMNYE-I-FSIPQYYKKGIIRKPDISQRLVYQF 331
           :||| || |:::: ||||||||::| || | |: |    .|  |.|
Db     352 DEVARRWGERQSKPNMNYD-KLSRALPYYYLKNIMIKVH-GKRYAYKF 397
```

RESULT 3
S47616
transcription factor elg - fruit fly (Drosophila melanogaster)
N;Alternate names: transcription factor ets-97D
C;Species: Drosophila melanogaster
C;Date: 13-Jan-1995 #sequence_revision 23-May-1997 #text_change 18-Sep-1998
C;Accession: S47616; S28822
R;Thos, S.M.; Xie, X.; Fujita, T.; Papas, T.S.; Watson, D.K.; Schulz, R.A.
Oncogene 7, 2471-2478, 1992
A;Title: Molecular characterization and structural organization of D-elg, an ets proto-o
A;Reference number: S47616; MUID:93096481
A;Accession: S47616
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-464 <THE>
A;Cross-references: EMBL:X68259; NID:g7942; PID:g7943
R;Chen, L.; Bunting, M.; Karim, F.D.; Thummel, C.S.
Dev. Biol. 151, 176-191, 1992
A;Title: Isolation and characterization of five Drosophila genes that encode an ets-rela
A;Reference number: S28819; MUID:92249640
A;Accession: S28822
A;Molecule type: mRNA
A;Residues: 298-449 <CHE>
A;Cross-references: EMBL:M88471; NID:g157189; PID:g552088
A;Note: the authors translated the codon AGC for residue 302 as Thr
C;Genetics:
A;Gene: FlyBase:Ets97D
A;Cross-references: FlyBase:FBgn0004510
A;Introns: 22/1; 60/2; 83/3; 322/1
C;Superfamily: transcription factor elg; ets DNA-binding domain homology; ets RII reg
C;Keywords: DNA binding; nucleus; transcription factor
F;190-263/Domain: ets RII regulatory region homology <EPP>
F;348-426/Domain: ets DNA-binding domain homology <ETS>

Query Match 16.5%; Score 294; DB 1; Length 464;
Best Local Similarity 29.1%; Pred. No. 7.5e-17;
Matches 92; Conservative 52; Mismatches 116; Indels 56; Gaps 12;

```
Qy      64 DMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPAGSLDLVPGGLTLEEHSLEQVQ 123
           |:| | ::: || |     :  |||   |    | :    :     :  |:   | ||:
Db     118 DVLKPTEAALAALAEEVVGQLSPPETASQ----KSSSSESPIKTPLKPMHKEDSEEESVE 173

Qy     124 SMVVGEVLK-DIETACK----LLNITADPMDWSPSNVQKWLLWTEHQYRLPPMGKAFQEL 178
             |  ||   ::  |     | |     :|: ::|  || |   |: |  :  :  ::
Db     174 GKDVKPVLNWVLDSKFKREQIRLKIPEAANEWTHAHVTYWLEWAVKQFELVGINMSDWQM 233

Qy     179 AGKELCAMSEEQFRQRSPLG-GDVLHAHLDIWKSAAWM-------KERTSP--------- 222
           |:|||||: |:| |: |   |::   || : |    ::     :|:  |
Db     234 NGQELCAMTHEEFNQKLPRDPGNIFWTHLQLLKECNFVSVVHKRAEEQRKPKQPRIMSAN 293

Qy     222 -------GAI------------HYCASISEESWIDS-------EVDSSCSGQPIHLWQFL 255
                  |::             :| | || | |         : |  :|| : |||||
Db     294 SISTNSGGSLSLEQPIMPKSYQSVKSSDSVESTTSSMNPSNYTTIGSGNNGQ-VQLWQFL 352

Qy     256 KELLLKPHSYGRFIRWLNKEKGIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIPQYYKK 315
           | :|   :  | |:  | | ||: |  :|||||| :|||:|||||:|||||::| ||
Db     353 LE-ILTDCEHTDVIEWVGTE-GEFKLTDPDRVARLWGFKKNKPAMNYEKISRALRYYYDG 410

Qy     316 GIIRKPDISQRLVYQF 331
           :| |    :|  |:|
Db     411 DMISKVS-GKRFAYKF 425

RESULT 4
I38739
nuclear respiratory factor-2 subunit alpha - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 18-Dec-1998
C;Accession: I38739
R;Gugneja, S., Virbasius, J.V., Scarpulla, R.C.
Mol. Cell. Biol. 15, 102-111, 1995
A;Title: Four structurally distinct, non-DNA-binding subunits of human nuclear respir
A;Reference number: I38739; MUID:95097980
A;Accession: I38739
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-454 <RES>
A;Cross-references: EMBL:U13044; NID:g521892; PID:g531893
C;Superfamily: transcription factor elg; ets DNA-binding domain homology; ets RII reg
F;174-245/Domain: ets RII regulatory region homology <EPP>
F;322-400/Domain: ets DNA-binding domain homology <ETS>

Query Match 16.4%; Score 293; DB 2; Length 454;
Best Local Similarity 28.1%; Pred. No. 8.8e-17;
Matches 86; Conservative 49; Mismatches 125; Indels 46; Gaps 9;

```
Qy   60 LSYFDMLYPEDSSWAAKAPGASSPEEPPEEPEQCPVIDSQAPAGSLDLVPGGLTLEEHSL 119
           |: ::: | |: | | : |: || || |: : :
Db   108 LNILEIVKPADTVEVVIDPDAHHAESEAHLVEEAQVI-------TLDGTKHITTISDETS 160

Qy   120 EQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLLWIEHQYRLPPMGKAFQELA 179
           ||| |:   : | | ||: || | |:::| :: : :   ::
Db   161 EQVTRWAA--ALEGYRKEQERLGIPYDPIQWSTDQVLHWVVWVMKEFSMTDIDLITLNIS 218
```
```

```
Qy      19 GKELCAMSEEQFPQPSPLGGDVLHAHLDIWKSAAWMKER---------------------- 219
           :::|::::| | :|  :  :::| :|||:: :      :
Db     214 GPELCSLNQEDFFQRVP PGEIIWSHLELLRKYVLASQEQQMNEIVTIDQPVQIIPASVQ 273

Qy     220 ----TSPGAIHYCASTS--EESWTDSEVDSSCSGQ------PIHLWQFLKELLLKPHSYG 266
               :  |: ||  :  : :  |  | |  |          | ||||| |||    :
Db     274 SAIPIIKVINSGAKAAKVQPAPPISGEDRSSPGNPTGNNGQIQLWQFLLELLTDKDARD 333

Qy     267 PFIPWLNKEEKGIFKIEDSAQVARLWGIPKNPPAMNYDKLSRSIPQYYKKGIIPKPDISQR 326
           |  :  : | ||:     ||: || |||:| |||.||||..| ||   .| |    .|
Db     334 CISWVGDE-GEFKLNQPELVAQKWGQPKNKPIMNYEKLSRALRYYYDGDMIKKVQ-GKR 391

Qy     327 LVYQFV 332
           |::|
Db     392 PVYKPV 400


RESULT   5
S60754
transcription factor erg - chicken
C;Species: Gallus gallus (chicken)
C;Date: 27-Apr-1996 #sequence_revision 23-May-1997 #text_change 10-Jul-1998
C;Accession: S60754
R;Dhordain, P., Dewitte, F., Desbiens, X., Stehelin, D., Duterque-Coquillaud, M.
Mech. Dev. 50, 17-28, 1995
A;Title: Mesodermal expression of the chicken erg gene associated with precartilaginous
A;Reference number: S60754; MUID:95329425
A;Accession: S60754
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-478 <DHO>
A;Cross-references: EMBL:X77159; NID:g790439; PID:g790440
C;Genetics:
A;Gene: erg
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: DNA binding; phosphoprotein; proto-oncogene; transcription factor; transform
F;1-84/Domain: ets RII regulatory region homology <ETS2>
F;312-390/Domain: ets DNA-binding domain homology <ETS>


  Query Match             16.4%;  Score 292;  DB 1;  Length 478;
  Best Local Similarity   26.9%;  Pred. No. 1.1e-16;
  Matches   99;  Conservative   39;  Mismatches  113;  Indels  114;  Gaps   14;

Qy      67 YPEDSSWAAKAPGASSREEPP----------------EEPEQCPVIDSQAPAGSLDLV 108
           :|  : : :      :||                    |: | |       | | |
Db      41 YSLTPKMSPSVPSQDWLSLFPAPVTIKMECNPNQVNGSRNSPDDCSVAKGGKMVSSSDNV 100

Qy     109 P---GGLILEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLLWTEHQ 165
           |   |:|     :|   |  |       : : : ||| || :|:||| |   :
Db     101 GMNYGSYMEEKHI PPPNMTTNE---------RRVIVPADPTLWSTDHVPQWLEWAVKE 149

Qy     166 GRIPMSDAAPLRIASKRIAMSEEQFPQPSP-LGGDVLHAHL--------------DIW 209
           | |: :   ||: : ||||| :::: |:: ::    |:| :||              |:
Db     150 YSIPNVDIIHDLKDGKFLCKMTKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDDVD 209

Qy     210 K----SAAWMKERPGPSPGAIHYCASTS----------------------EESW--------- 236
               :  | : : |:     :||                          :|
Db     210 KAIQNSPRMHARNTGGATFIFPNTSVYPEATQRITTRPDLPYEQAPRSAWTSHSHPTQS 269

Qy     237 --------------------------TDSEVDSSCSGQPIHLWQFLKELLLKPHSYGP 267
                                       | | : :  ||| | ||||| | ||   |
Db     270 KAIQPSSSTVPKIHTQRPQGIDPYQILGPTSSRLANPGSGQ-IQLWQFLLE-LLSDSSNSN 327

Qy     268 PFIPWLNKEEKGIFKIEDSAQVARLWGIPKNPPAMNYDKLSRSIPQYYKKGIIPKPDISQRL 327
           |  :     | ||: |  :||| || ||..| ||||||||::| || | |. :    .|
Db     328 CITWGEDTVGEFPMQTPEPVARPWQEEFSKPNMNYDKLSRALRYYYDKNIMTKVH-GKRY 385

Qy     328 VYQF 331
           |:|
Db     386 AYKP 389


RESULT   6
A48146
GA-binding protein alpha chain - human
N;Alternate names: GABPA; nuclear respiratory factor-2 alpha chain; transcription fac
C;Species: Homo sapiens (man)
C;Date: 21-Jan-1994 #sequence_revision 23-May-1997 #text_change 26-Feb-1999
C;Accession: A48146; A46303; A55903
R;Watanabe, H.; Sawada, J.; Yano, K.; Yamaguchi, K.; Goto, M.; Handa, H.
Mol. Cell. Biol. 13, 1385-1391, 1993
A;Title: cDNA cloning of transcription factor E4TF1 subunits with Ets and notch motif
A;Reference number: A48146; MUID:93180783
A;Accession: A48146
A;Status: not compared with conceptual translation
A;Molecule type: mRNA; protein
A;Residues: 1-454 <WAT>
A;Note: sequence extracted from NCBI backbone (NCBIP:125762)
A;Note: parts of this sequence were determined by protein sequencing
R;Virbasius, J.V.; Virbasius, C.A.; Scarpulla, R.C.
Genes Dev. 7, 380-392, 1993
A;Title: Identity of GABP with NRF-2, a multisubunit activator of cytochrome oxidase
A;Reference number: A46303; MUID:93194058
A;Accession: A46303
A;Molecule type: protein
A;Residues: 237-248,429-448 <VIR>
R;Cardot, P., Pastier, D., Lacorte, J.M., Mangeney, M., Zannis, V.I.; Chambar, J.
Biochemistry 33, 12139-12148, 1994
A;Title: Purification and characterization of nuclear factors binding to the negative
ets-related protein.
A;Reference number: A55903; MUID:95001929
A;Accession: A55903
A;Molecule type: protein
A;Residues: 26-34,'A',36,'X',38,'X',40-46,161-165,167-174,356-359 <CAR>
C;Genetics:
A;Gene: GDB:GABPA; E4TF1-60; E4TF1A
A;Cross-references: GDB:138476; OMIM:600609
A;Map position: 21q21-21q22.1
C;Complex: GA-binding protein is a heterotetramer of two alpha and two beta-type chai
C;Function:
A;Description: a transcription factor that binds (via the alpha chain) to GA-rich pro
A;Pathway: known to promote transcription of apolipoprotein A-II, cytochrom c oxidase
C;Superfamily: transcription factor elg; ets DNA-binding domain homology; ets RII reg
C;Keywords: DNA binding; nucleus; transcription factor
F;174-245/Domain: ets RII regulatory region homology <ETS>
F;322-400/Domain: ets DNA-binding domain homology <ETS>


  Query Match             16.2%;  Score 289;  DB 1;  Length 454;
  Best Local Similarity   27.8%;  Pred. No. 1.9e-16;
  Matches   85;  Conservative   49;  Mismatches  126;  Indels   46;  Gaps    9;

Qy      60 LSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPAGSLDLVPGGLTLEEHSL 119
           |:  ::: | |:      | |   |       |:  ||       :||      |: : :
Db     108 LNILEIVKPADIVEVVIDPDAHHAESEAHLVEEAQVI-------ILLGIKHIIIISDEIS 160

Qy     120 EQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLLWTEHQYPLPPMGKAPQELA 179
           |||       |:      : | |  || || ||   |  |::|   :: :  :      ::
Db     161 EQVTPWAA--ALEGYRKEQEPLGIFYDPIQWSIDQVLHWVVWVMEEFSMTDIDLTTLNIS 218

Qy     180 GKELCAMSEEQFPQPSPLGGDVLHAHLDIWKSAAWMKER---------------------- 219
           |:|||::..| | || |  |..| .||.. :    :
Db     219 GPELCSLNQEDFFQRVP-PGEIIWSHLELLRKYVLASQEQQMNEIVTIDQPVQIIPASVQ 277

Qy     219 -TSPGAIHYCASTSEESWTD-----SEVDSSCSGQ------PIHLWQFLKELLLKPHSYG 266
            :|  |    |::: :          |  | |  |          | ||||| :||    :
Db     278 SAIPIIIKVINSSAKAAKVQPAPPISGEDRSSPGNPTGNNGQIQLWQFLLTLLTDKDARD 337

Qy     267 PFIPWLNKEEKGIFKIEDSAQVARIWGIPKNPPAMNYDKLSRSIPQYYKKGIIPKPDISQR 326
           |  |:  | | ||:     ||: || |||:| ||||||||::| ||   :|  :    :|
Db     338 -CISWVGDE-GEFKLNQPELVAQKWGQPKNKPIMNYEKLSRALRYYYDGDMIKKVQ-GKR 394

Qy     327 LVYQFV 332
           ||:||
```

```
Db      995 FVYKFV 400


RESULT   7
A40858
GA-binding protein alpha chain - mouse
N;Alternate names: GABPA; nuclear respiratory factor-2 alpha chain; transcription factor
C;Species: Mus musculus (house mouse)
C;Date: 28-Feb-1992 #sequence_revision 23-May-1997 #text_change 10-Jul-1998
C;Accession: A40858
R;LaMarco, K.; Thompson, C.C.; Byers, B.P.; Walton, E.M.; McKnight, S.L.
Science 253, 789-792, 1991
A;Title: Identification of ets- and notch-related subunits in GA binding protein.
A;Reference number: A40858; MUID:91343912
A;Accession: A40858
A;Molecule type: mRNA; protein
A;Residues: 1-454 <LAM>
A;Cross-references: GB:M74515; NID:g193382; PID:g193383
A;Note: parts of this sequence were determined by protein sequencing
C;Genetics:
A;Gene: Gabpa
A;Map position: 16 47.0
C;Complex: GA binding protein is a heterotetramer of two alpha and two beta-type chains.
C;Function:
A;Description: a transcription factor that binds (via the alpha chain) to GA-rich promot
A;Pathway: known to promote transcription of apolipoprotein A-II, cytochrom c oxidase ch
C;Superfamily: transcription factor elg; ets DNA-binding domain homology; ets RII regula
C;Keywords: DNA binding; nucleus; transcription factor
F;174-245/Domain: ets RII regulatory region homology <ERR>
F;322-405/Domain: ets DNA-binding domain homology <ETS>


  Query Match             15.5%;  Score 277;  DB 1;  Length 454;
  Best Local Similarity   28.8%;  Pred. No. 1.9e-15;
  Matches   84;  Conservative   43;  Mismatches  116;  Indels   46;  Gaps    9;

Qy      78 PE PGASSREEPPEEPEQCPVIDSQAPAGSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETA 137
           |  |    |       |:  ||       :||      |: : : |||        |:
Db     126 PDAHHAEAEAHLVEEAQVI-------TLDGTKHITTISDETSEQVTPWAA--ALEGYPKE 176

Qy     138 PKIGNIIAIPMPWSPGNVQKWLLWIEHQYRLPPMGKAPQELAGKELCAMSEDQFFQRSPL 197
           : | |  ||: ||   |  |::|   :: : :      ::|:|||:::: | || |
Db     177 QERLGIPYDPIRWSTDQVLHWVVWVMKEFSMTDIDLTTLNISGRELCSLNQEDFFQRVP- 236

Qy     198 GGDVGHAHLDIWKSAAWMKER----------------------TSPGAIHYCASTS---- 232
             |::| :|||: :      :                      :|  |    |::
Db     237 RGEIIWSHLELLFKYVLASQEQQMNEIVTIDQPVQIIPASVPPATPTTIKVINSSAKAAK 295

Qy     233 -EESWTDSEVDSSCSGQ------PIHLWQFLKELLLKPHSYGRFIRWLNKEKGIFKIEDS 284
           : |  |  | |  |         | |||||| |||    :  : |:|: | | ||:
Db     296 VQKSPRISGEDRSSPQNETQNNGQIQLWQFLLELLTDKDARD-CISWVGDE-GEFKLNQF 353

Qy     285 A VAALAWTIEANFIAMNYLKLSKSIFQYYKKGIIPKPDISQRLVYQFV 332
           *: || |||:| |||:|||||:| ||   :| |    :| ||:||
Db     354 EIVAQEWTLEFNEFIMNYIKLSPALRYYYDGDMICKVQ-GKRFVYKFV 400


RESULT   8
S49013
transforming protein fli   African clawed frog
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Xenopus laevis (African clawed frog)
C;Date: 07-May-1995 #sequence_revision 24-May-1997 #text_change 05-Sep-1997
C;Accession: S49013
R;Meyer, D.; Wolff, C.M.; Stiegler, P.; Senan, F.; Befort, N.; Befort, J.J.; Remy, P.
Mech. Dev. 44, 109-121, 1993
A;Title: Xl fli, the Xenopus homologue of the fli-1 gene, is expressed during embryogene
A;Reference number: S49013; MUID:94206844
A;Accession: S49013
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-454 <MEY>
A;Cross-references: EMBL:X66979; NID:g505486; PID:g505487
C;Genetics:
A;Gene: fli
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII reg
C;Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transforming
F;117-191/Domain: ets RII regulatory region homology <ETS2>
F;284-362/Domain: ets DNA-binding domain homology <ETS>


  Query Match             15.4%;  Score 275;  DB 1;  Length 453;
  Best Local Similarity   32.4%;  Pred. No. 2.7e-15;
  Matches   78;  Conservative   29;  Mismatches   78;  Indels   56;  Gaps    9;

Qy     143 ITADPMDWSPSNVQKWLLWTEHQYRLPPMG-KAPQELAGKELCAMSEDQF-RQRSPLGGD 200
           : |||  ||  :|::||  |   :| |  :    || : |||||  ||:| |  |  |    :
Db     125 VPADPALWSQDHVRQWLEWAIKEYGLVEIDCSLFQNIDGKELCKMSKEDFLRSISIYNTE 184

Qy     201 VLHAHLDIWKSAA---------------------------WMPERTSP-------- 222
           || :||:  : ::                           |    :||
Db     185 VLLSHLNYLPDSSSSLGYNTQAHTDQSSRLTAKEDPSYEAVRPSGWGNSMSSPVTKSPPM 244

Qy     222 GAIHYCASTSEESW-----------TDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIR 270
           |       : ::              | | :  ||| | ||||| | ||   |    |
Db     245 GGTQNVNKSGDQQRSQPDPYQILGPTSSRLANPGSGQ-IQLWQFLIE-LLSDSSNASCIT 302

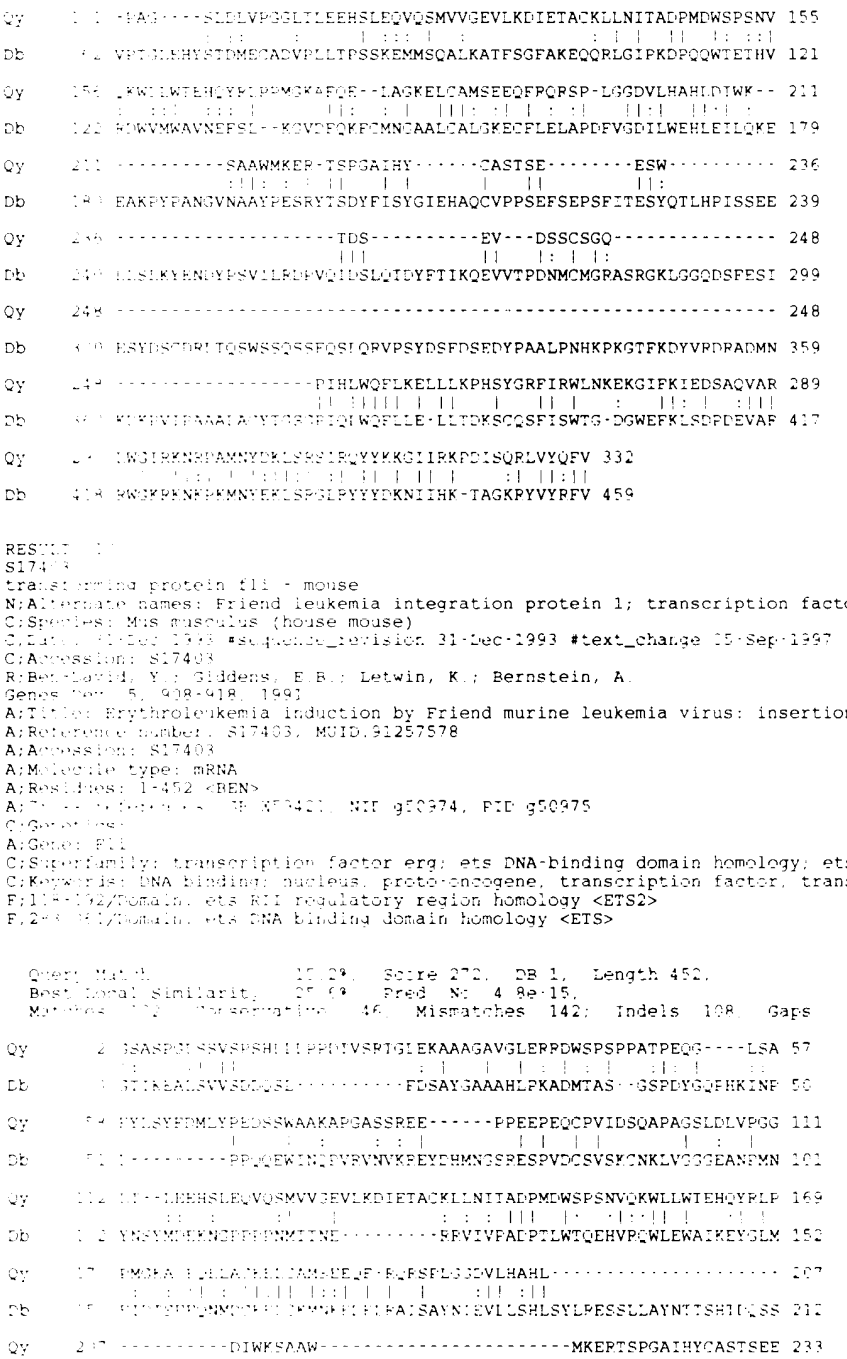
Qy     271 WLNKEKGIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQ 330
           |     | ||: |  :||| || ||::| ||||||||::| || | |: |    :|  |:
Db     303 WEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKSIMTKVH-GKRYAYK 360

Qy     331 F 331
           |
Db     361 F 361


RESULT   9
TVCHET
transcription factor ets-1, splice form b - chicken
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 30-May-1997
C;Accession: A28875; A28610
R;Leprince, D.; Duterque-Coquillaud, M.; Li, R.P.; Henry, C.; Flourens, A.; Debuire,
J. Virol. 62, 3233-3241, 1988
A;Title: Alternative splicing within the chicken c-ets-1 locus: implications for tran
A;Reference number: A28875; MUID:88300873
A;Accession: A28875
A;Molecule type: mRNA
A;Residues: 1-485 <LEP>
A;Cross-references: GB:M21823; NID:g211465; PID:g211466
R;Watson, D.K.; McWilliams, M.J.; Papas, T.S.
Virology 164, 99-105, 1988
A;Title: Molecular organization of the chicken ets locus.
A;Reference number: A28610; MUID:88206088
A;Accession: A28610
A;Molecule type: DNA
A;Residues: 1-485 <WAT>
A;Cross-references: GB:M20507; GB:M20508; GB:M20509; GB:M20510; GB:M20511; GB:M20512;
C;Genetics:
A;Gene: ets-1
C;Superfamily: transcription factor ets; ets DNA-binding domain homology; ets RII req
C;Keywords: alternative splicing; DNA binding; nucleus; proto-oncogene; transcription
F;101-174/Domain: ets RII regulatory region homology <ETS2>
F;141-174/Region: helix-loop-helix #status predicted
F;381-459/Domain: ets DNA-binding domain homology <ETS>
F;421-427/Region: nuclear location signal


  Query Match             15.2%;  Score 272;  DB 1;  Length 485;
  Best Local Similarity   24.2%;  Pred. No. 5.3e-15;
  Matches  112;  Conservative   42;  Mismatches  109;  Indels  200;  Gaps   18;

Qy      65 MLYPEDSSWAAKAP----------GASSREEPPEEP----------EQCPVIDSQA--- 101
           | |  |::  :  |          |||| | |  |           ||  |  |
Db       2 MSYYMDITIGSTGPYPLARPGVMQGASSCCEDPWMPCRLQSACCPPRSCCPPWDEAAIQE 61
```



RESULT 11
S17403
transforming protein fli - mouse
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Sep-1997
C;Accession: S17403
R;Ben-David, Y.; Giddens, E.B.; Letwin, K.; Bernstein, A.
Genes Dev. 5, 908-918, 1991
A;Title: Erythroleukemia induction by Friend murine leukemia virus: insertional activati
A;Reference number: S17403; MUID:91257578
A;Accession: S17403
A;Molecule type: mRNA
A;Residues: 1-452 <BEN>
A;Cross-references: GB:X57421; NID:g50974; PID:g50975
C;Genetics:
A;Gene: FLI
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: DNA binding; nucleus; proto-oncogene; transcription factor; transforming pro
F;1-192/Domain: ets RII regulatory region homology <ETS2>
F;243-331/Domain: ets DNA binding domain homology <ETS>

RESULT 11
I37565
transforming protein fli, long splice form - human
N;Alternate names: Friend leukemia integration protein 1; transcription factor ERGB
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1996 #sequence_revision 30-May-1997 #text_change 20-Mar-1998
C;Accession: I37565; S29843; S35506; A49000; A49015; I54170
R;Delattre, O.; Zucman, J.; Plougastel, B.; Desmaze, C.; Melot, T.; Peter, M.; Kovar,
Nature 3599, 162-165, 1992
A;Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocatio
A;Reference number: I37378; MUID:92396239
A;Accession: I37565
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-452 <RES>
A;Cross-references: EMBL:X67001; NID:g32529; PID:g32530
R;Hromas, R.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Klems
Biochim. Biophys. Acta 1172, 155-158, 1993
A;Title: Human FLI-1 localizes to chromosome 11Q24 and has an aberrant transcript in
A;Reference number: S29843; MUID:93176799
A;Accession: S29843
A;Molecule type: mRNA
A;Residues: 1-68,'V',70-129,'A',131-132,'V',134-322,'Q',324-425,427-452 <HRO1>
A;Cross-references: EMBL:M93255
A;Note: the authors translated the codon GTG for residue 69 as Glu
R;Hromas, R.A.; May, W.; Denny, C.; Raskind, W.; Moore, J.; Maki, R.A.; Beck, E.; Kle
submitted to the EMBL Data Library, May 1993
A;Description: Human FLI-1, an ETS oncogene family member preferentially expressed in
A;Reference number: S35506
A;Accession: S35506
A;Molecule type: mRNA
A;Residues: 1-68,'V',70-76,78-129,'A',131-132,'V',134-322,'Q',324-425,427-452 <HRO2>
A;Cross-references: EMBL:M93255; NID:g182659; PID:g182660
R;Prasad, D.D.; Rao, V.N.; Reddy, E.S.
Cancer Res. 52, 5833-5837, 1992
A;Title: Structure and expression of human Fli-1 gene.
A;Reference number: A49000; MUID:93007976
A;Accession: A49000
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-322,'Q',324-452 <PRA>
A;Cross-references: GB:S45205; NID:g257353; PID:g257354
A;Note: sequence extracted from NCBI backbone (NCBIN:115336, NCBIP:115337)
R;Watson, D.K.; Smyth, F.E.; Thompson, D.M.; Cheng, J.Q.; Testa, J.R.; Papas, T.S.; S
Cell Growth Differ. 3, 705-713, 1992
A;Title: The ERGB/Fli-1 gene: isolation and characterization of a new member of the f
A;Reference number: A49015; MUID:93075640
A;Accession: A49015
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-390,'Q',392-425,427-452 <WAI>
A;Cross-references: GB:M98833; NID:g182188; PID:g182189
A;Note: sequence extracted from NCBI backbone (NCBIN:119390, NCBIP:119392)
R;Raud, V.; Lipinski, M.; Passart, E.; Poliquin, L.; Bergeron, D.
Genomics 11, 223-224, 1991
A;Title: The human homolog of the mouse common viral integration region, FLI1, maps t
A;Reference number: I54170; MUID:92112219
A;Accession: I54170
A;Status: preliminary; nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6 <PRE2>

A;Cross-references: GB:L47616; NID:g1000863; PID:g1000864
C;Genetics:
A;Gene: HUB:FLI1
A;Cross-references: GDB:127565; OMIM:193067
A;Map position: 11q24.1-11q24.3
A;Introns: 77/3
C;Superfamily: transcription factor erg; ets DNA-binding domain homology; ets RII regula
C;Keywords: alternative splicing; DNA binding; nucleus, proto-oncogene, transcription fa
F;119-192/Domain: ets RII regulatory region homology <ETS2>
F;283-361/Domain: ets DNA binding domain homology <ETS>

Query Match 15.1%; Score 270; DB 1; Length 452;
Best Local Similarity 25.9%; Pred. No. 7e-15;
Matches 103; Conservative 50; Mismatches 137; Indels 108; Gaps 17;

Qy 2 GSASPGLSSVSPSHLLLPPDIVSRIGLEKAAAGAVGLEPRDWSPSPPATPEQG----LSA 57
Db 3 GTIKEALSVVSDDQSL----------FDSAYGAAAHLPKADMTAS--GSPDYGQPHKINP 50

Qy 58 PYLSYFDMLYPEDSSWAAKAPGASSREE------PPEEPEQCPVIDSQAPAGSLDLVPGG 111
Db 51 L--------PPQQEWINQPVRVNVKREYDHMNGSRESPVDCSV------SKCSKLVGGG 95

Qy 112 LL--------LEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLLWTE 163
Db 96 ESNFMNYNSYMDEKNSPPPNMTTNE---RRVIVPADPTLWTQEHVRQWLEWAI 146

Qy 164 HQYRLPPMGKA-FQELAGKELCAMSEEQFRQRSPL-GGDVLHAHL---------- 207
Db 147 KEYSLMPIDSEFQNMEGKFIFKMNKFDFLPATILYNTEVLLSHLSYLRESSLLAYNTTS 206

Qy 208 ----------DIWKSAAWMKERTSP-------GAIHYCASTSEE---------- 234
Db 207 HIDQSSPLSVKEDPSYDSVRRGAWGNNMNSGLNKSPPLGGAQTISKNTEQRPQPDPYQIL 266

Qy 234 SWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEKGIFKIEDSAQVARLWGI 293
Db 267 GPTSSRLANPGSGQ-IQLWQFLLE-LLSDSANASCITW-EGTNGEFKMTDPDEVARRWGE 323

Qy 294 PENPKAMNYDKLSRSLRYYYKKGIIPKPDISQRLVYQF 331
Db 324 PKKPNMNYDKLSRALRYYYDKNIMTKVH-GKRYAYKF 360

RESULT 12
TVFVES
transcription factor ets - avian erythroblastosis virus (strain E26)
C;Species: avian erythroblastosis virus
C;Date: 01-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 30-May-1997
C;Accession: B01448; A91448
R;Nunn, M.F.; Seeburg, P.H.; Moscovici, C.; Duesberg, P.H.
Nature 308, 791-795, 1984
A;Title: Tripartite structure of the avian erythroblastosis virus E26 transforming gene.
A;Reference number: A91448; MUID:84068165
A;Accession: B01448
A;Molecule type: DNA
A;Residues: 1-488 <NUN>
A;Cross-references: GB:X00144
C;Comment: This protein is translated as a myb-ets polyprotein; only the ets part is sho
C;Genetics:
A;Gene: ets
C;Superfamily: transcription factor ets; ets DNA-binding domain homology; ets RII regula
C;Keywords: DNA binding; oncogene; transforming protein
F;107-174/Domain: ets RII regulatory region homology <ETS2>
F;141-174/Region: helix loop helix #status predicted
F;281-459/Domain: ets DNA binding domain homology <ETS>
F;421-427/Region: nuclear location signal

Query Match 14.0%; Score 256; DB 1; Length 488;
Best Local Similarity 23.8%; Pred. No. 1.7e-14;
Matches 110; Conservative 44; Mismatches 109; Indels 202; Gaps 18;

Qy 65 MLYPEDSSWAAKAP----------GASSREEPPEEP-----------EQCPVIDSQA--- 101
Db 2 MSYYMDTTIGSTGPYPLARPGVMQGASSCTFDPWMPCPLQSACPPPPSCCPPWDEAAIQE 61

Qy 101 -PAG----SLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNV 155
Db 62 VPTGLEHYSTDMECADVPLLTPSSKEMMSQALKATFSGFAKFQQPLGIPKDPQQWTETHV 121

Qy 156 QKWLLWIEHQYRLPPMGKAFQE--LAGKELCAMSEEQFRQRSP-LGGDVLHAHLDIWK-- 211
Db 122 RDWVMWAVNEFSL--KGVDFQKFCMNGAALCALGKECFLELAPDFVGDILWEHLEILQKE 179

Qy 211 ----------SAAWMKER-TSPGAIHY------CASTSE--------ESW-- ------- 236
Db 180 EAKPYPANGVNAAYPESRYTSDYFISYGIEHAQCVPPSEFSEPSFITESYQTLHPISSEE 239

Qy 236 --------------------TDS----------EV---DSSCSGQ--------------- 248
Db 240 LLSLKYENDYPSVILRDPVQTDSLQTDYFTIKQEVVTPDNMCMGKVSPGKLGGQDSFESI 299

Qy 248 -- 248
Db 300 ESYDSCDRLTQSWSSQSSFQSLQPVPSYDSFDSEDYPAALPNHKPKGTFKDYVRDRADMN 359

Qy 248 -----------------PIHLWQFLKELLLKPHSYGRFIRWLNKEKGIFKIEDSAQVAR 289
Db 360 KDKPVIPAAALAGYTGSGPIQLWQFLLE-LLTDKSCQSFISWTG-DGWEFKLSDPDEVAR 417

Qy 290 LWGIRKNRPAMNYDKLSRSIRQYYKKGIIPKPDISQRLVYQFV 332
Db 418 RWGKRKNKPKMDYEKLSRGLRYYYDKNVIHK-TAGKRYVYRFV 459

RESULT 13
G02318
Ets transcription factor - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C;Accession: G02318
R;Libermann, T.
submitted to the EMBL Data Library, December 1995
A;Reference number: H01044
A;Accession: G02318
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-581 <LIB>
A;Cross-references: EMBL:U43188; NID:g1420888; PID:g1420889
C;Genetics:
A;Gene: NERF-2
C;Superfamily: ets DNA-binding domain homology
F;198-278/Domain: ets DNA-binding domain homology <ETS>

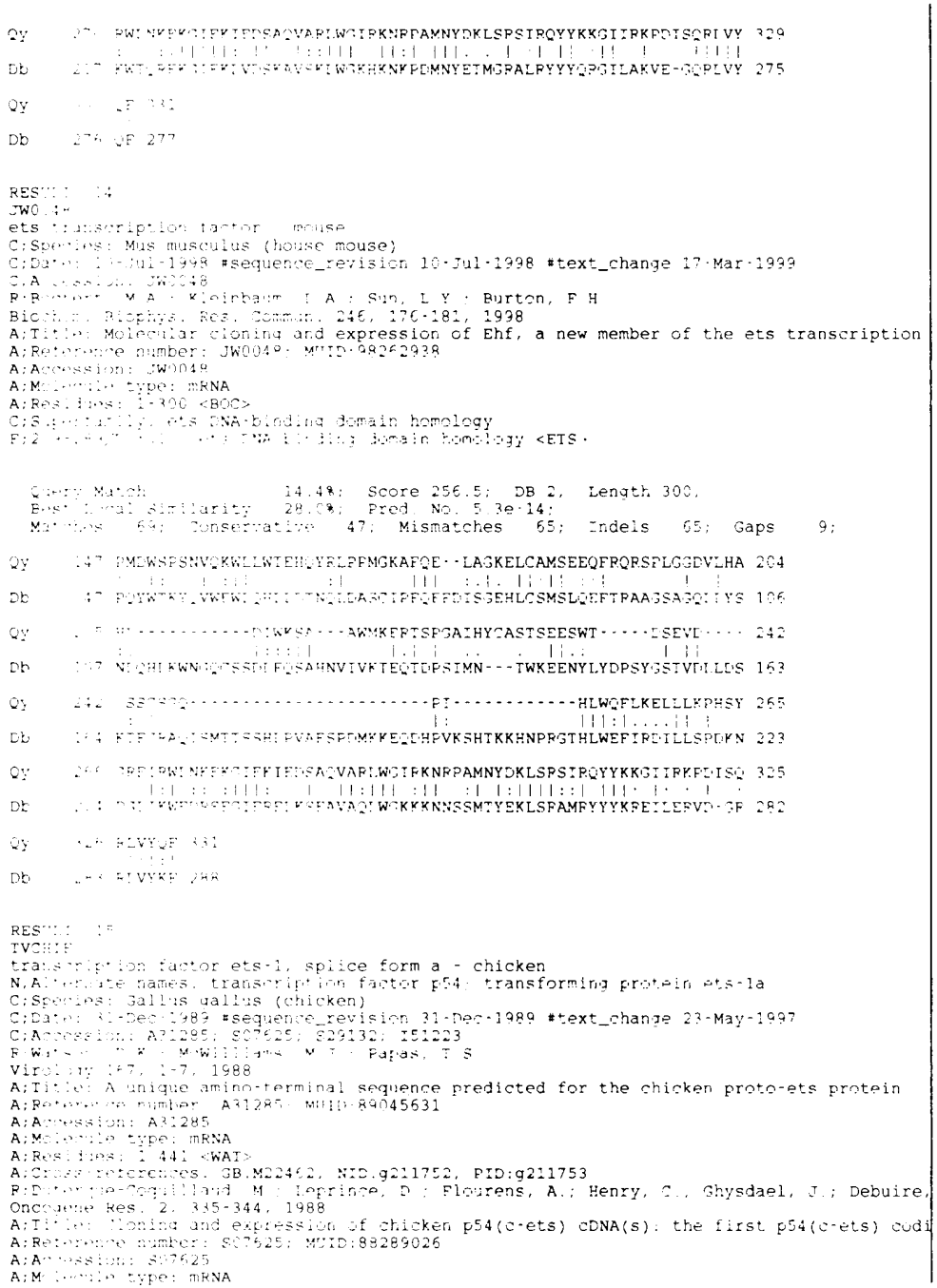
Query Match 14.4%; Score 257; DB 2; Length 581;
Best Local Similarity 25.5%; Pred. No. 1.2e-13;
Matches 77; Conservative 50; Mismatches 97; Indels 78; Gaps 9;

Qy 45 PSPPATPEQGLSAFYLSYFDMLY-PEDSSWAAKAPGASSREEPPEEPEQCPVIDS----- 99
Db 39 PVPSARLEQGYAAQVLVYDDETYMMQDVA----------EEQEVETENVETVEASVHSS 87

Qy 99 ---------QAPAGSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMD 149
Db 88 NAHCTDKTIEAAEALLHMESPTCLRDSRSPEFIHAAMRPDVI--TETVVEVSTEESEPMD 145

Qy 150 WSPSNVQKWLLWTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIW 209
Db 146 TSP--------------IPTSPDSHEPMKKKK---VGRKPKTQQSPISNG--------- 179

Qy 210 KSAAWMKERISPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFIKFILLKPHSYGRFI 269
Db 179 SPEIGIKKKPPEG---------- -----KQNTTYIWFFIDLLQPKNTCPRYI 216



```
Qy     [illegible] RWINKPKGIPKIPDSAQVARLWGIPKNRPAMNYDKLSPSIRQYYKKGIIRKPDISQRIVV 329
Db     217 FWTLPPPIPPKIVDSPAVSPIWGKHKNKPDMNYETMGPALPYYYQPGILAKVE-GQPLVV 275

Qy     [illegible] QF 341
Db     276 QF 277

RESULT  14
JW0049
ets transcription factor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 14-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 17-Mar-1999
C;Accession: JW0049
R;Bhat, N. A.; Klotzbaum, I. A.; Sun, L. Y.; Burton, F. H.
Biochem. Biophys. Res. Commun. 246, 176-181, 1998
A;Title: Molecular cloning and expression of Ehf, a new member of the ets transcription
A;Reference number: JW0049; MUID:98262938
A;Accession: JW0049
A;Molecule type: mRNA
A;Residues: 1-300 <BOC>
C;Superfamily: ets DNA-binding domain homology
F;2-[illegible]/Domain: ets DNA binding domain homology <ETS>

  Query Match             14.4%;  Score 256.5;  DB 2,  Length 300;
  Best Local Similarity   28.0%;  Pred. No. 5.3e-14;
  Matches   64;  Conservative   47;  Mismatches   65;  Indels   65;  Gaps    9;

Qy     147 PMDWSPSNVQKWLLWIEHQYRLPPMGKAFQE--LAGKELCAMSEEQFRQRSPLGGDVLHA 204
Db      47 PQYWTKAY.VWFWLQNIITNQLDASCIPPQFPDISGEHLCSMSLQEFTPAAGSAGQIIYS 106

Qy     205 L-W-----------DIWKSA---AWMKEPTSPGAIHYCASTSEESWT-----ESEVD---- 242
Db     107 NIQHIKWNQQTSSDIFQSAHNVIVKTEQTDPSIMN---TWKEENYLYDPSYGSTVPLLDS 163

Qy     243 SSTSTQ----------------------PI------------HLWQFLKELLLKPHSY 265
Db     164 KTPTPAQISMTTSSHIPVAFSPDMKFEQDHPVKSHTKKHNPRGTHLWEFIRDILLSPDKN 223

Qy     266 RPFPWLNKFKGIFFIFDSAQVARLWGIRKNRPAMNYDKLSPSIRQYYKKGIIRKPDISQ 325
Db     224 PNIPWWPNQFPGIPPPLKGPAVAQIWGKKKNNSSMTYEKLSPAMRYYYKREILEPVD-GP 282

Qy     326 RLVYQF 331
Db     283 RIVYKF 288

RESULT  15
IVCHIP
transcription factor ets-1, splice form a - chicken
N;Alternate names: transcription factor p54; transforming protein ets-1a
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 23-May-1997
C;Accession: A31285; S07625; S29132; I51223
R;Watson, D. K.; McWilliams, M. J.; Papas, T. S.
Virology 167, 1-7, 1988
A;Title: A unique amino-terminal sequence predicted for the chicken proto-ets protein
A;Reference number: A31285; MUID:89045631
A;Accession: A31285
A;Molecule type: mRNA
A;Residues: 1-441 <WAT>
A;Cross-references: GB:M22462; NID:g211752; PID:g211753
R;Dutertre-Coquillard, M.; Leprince, D.; Flourens, A.; Henry, C.; Ghysdael, J.; Debuire,
Oncogene Res. 2, 335-344, 1988
A;Title: Cloning and expression of chicken p54(c-ets) cDNA(s): the first p54(c-ets) codi
A;Reference number: S07625; MUID:88289026
A;Accession: S07625
A;Molecule type: mRNA
A;Residues: 1-116,'R',118-441 <DUT>
A;Cross-references: EMBL:X13026; NID:g63179; PID:g63180
A;Note: the authors translated the codon AGC for residue 26 as Phe and CGG for residu
R;Chen, J.H.
Oncogene Res. 2, 371-384, 1988
A;Title: Complementary DNA clones of chicken proto-oncogene c-ets-1 sequence divergenc
A;Reference number: S29132; MUID:88289029
A;Accession: S29132
A;Molecule type: mRNA
A;Residues: 1-116,'R',118-441 <CHE>
A;Cross-references: EMBL:X13027; NID:g63382; PID:g63383
A;Experimental source: spleen
R;Crepieux, P.; Leprince, D.; Flourens, A.; Albagli, O.; Ferreira, F.; Stehelin, D.
Gene Expr. 3, 215-225, 1993
A;Title: The two functionally distinct amino termini of chicken c-ets-1 products aris
A;Reference number: I51223; MUID:94093304
A;Accession: I51223
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-22 <CRE>
A;Cross-references: GB:S68254; NID:g471777; PID:g471778
C;Genetics:
A;Gene: ets-1
C;Superfamily: transcription factor ets; ets DNA-binding domain homology; ets RII reg
C;Keywords: alternative splicing; DNA binding; nucleus; phosphoprotein; proto-oncogen
F;57-130/Domain: ets RII regulatory region homology <ETS2>
F;97-130/Region: helix-loop-helix #status predicted
F;337-415/Domain: ets DNA-binding domain homology <ETS>
F;377-383/Region: nuclear location signal

  Query Match             14.3%;  Score 254.5;  DB 1,  Length 441;
  Best Local Similarity   24.0%;  Pred. No. 1.3e-13;
  Matches   96;  Conservative   39;  Mismatches   88;  Indels  177;  Gaps   15;

Qy     105 LDLVF------GGLILEEHSLEQVQSMVVGEVLKQIFTACKLLNITAQFMDWSFSNVQKW 158
Db      21 IDLFPSPDMECADVPLLIPSSKEMMSQALKAIFSGFAKEQQPLGIPKDPQQWTETHVPDW 80

Qy     159 LLWTEHQYRLPPMGKAFQE--LAGKELCAMSEEQFRQRSP-LGGDVLHAHLDIWK----- 211
Db      81 VMWAVNEFSL--KGVDFQKFCMNGAALCALGKEQFLRLAPDFVGDIIWEHIEIQKFFAK 138

Qy     211 -------SAAWMKEP-TSPGAIHY------CASTSE--------ESW------------ 236
Db     139 PYPANGVNAAYPESRYTSIYFISYGIEHAQCVPPSEFSEPSFITESYQTLHFISSEELLS 198

Qy     236 -----------------IDS----------EV---DSSCSGQ----------------- 248
Db     199 LKYENDYPSVILRDPVQTDSLQIDYFTIKQEVVTPDNMCMGPASRGKLGGQDSFESIESY 258

Qy     248 ------------------------------------------------------------ 248
Db     259 DSCDPLTQSWSSQSSFQSLQPVPSYDSFDSFDYPAALPNHKPKGTFKDYVPDPADMNKDK 318

Qy     248 --------------PIHLWQFLKELLLKPHSYGPFIPWLNKFKGIFFIFDSAQVARLWG 292
Db     319 PVIPAAALAGYTGSGPIQLWQFLLE-LLIDKSCQSFISWIG-DGWFFKLSDPDFVARRWG 376

Qy     293 IRKNRPAMNYDKLSPSIRQYYKKGIIRKPDISQPLVYQFV 332
Db     377 KRKNKPKMNYEKLSRGLRYYYDKNIIHK-TAGKRYVYRFV 415

Search completed: November 20, 1999, 14:10:36
Job time: 23 sec
```